

SPATIO-TEMPORAL INFORMATION PROCESSING IN SINGLE NEURONS

A Dissertation

Presented to

the Faculty of the Department of Biology & Biochemistry

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Keith Kelleher

August, 2010

SPATIO-TEMPORAL INFORMATION PROCESSING IN SINGLE NEURONS

Keith Kelleher

APPROVED:

Dr. Krešimir Josić, Chairman

Dr. Peter Saggau, Co-Chairman
Baylor College of Medicine 77030

Dr. Michael Rea

Dr. Stuart Dryer

Dean, College of Natural Sciences and Mathematics

Acknowledgements

It is a pleasure to pay respects to those individuals who made my dissertation possible. I have been very lucky to have the opportunity to work with my two co-mentors, Dr. Peter Saggau, and Dr. Krešimir Josić. They have both repeatedly shown a remarkable ability to generate new ideas and strategies for experiments and data analysis. Their hard work and passion for advancing the course of science make them ideal role models for any graduate student.

I'd also like to thank Dr. Costa Colbert, who built much of the foundation for this project, through the experimental design, and through his vision of pyramidal neuron physiology. A very grateful thanks also for Dr. Yong Liang, who generated a lot of data for the back-propagation experiments, and who has always been a good friend, and supporter of my scientific efforts. Dr. Stuart Dryer and Dr. Michael Rea served as committee members and provided valuable feedback on the project.

Both Peter's and Kreso's labs held weekly lab meetings which provided a great venue for discussing the issues at hand. I truly appreciate the environment of the lab meetings, and the contributions of my fellow labmates. There have been many students and post-docs from whom I've absorbed both knowledge and enthusiasm.

Dr. Claudia-Patricia Alvarez-Baron provided advice and support for my scientific career, through discussion and inspiration, aside from her role as the sunshine of my life. I also thank my parents and family for their unwavering love and support for yeah, these many years.

SPATIO-TEMPORAL INFORMATION PROCESSING IN SINGLE NEURONS

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Biology & Biochemistry

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Keith Kelleher

August, 2010

Abstract

Learning in the brain is thought to be accomplished by changes in the connectivity between cells. There is evidence for regulation of the strength of existing synapses, as well as growth and pruning of synaptic contacts. One common learning rule, deemed Hebbian synaptic plasticity, declares that a synaptic connection between two cells will become stronger if the two cells are repeatedly co-activated. The back-propagating action potential, bAP, is thought to convey the necessary feedback to the synapses required for this coincidence detection.

We study the behavior of bAPs by initiating them from the soma, while measuring the degree of activation from many points on the dendrites, using the fluorescent calcium indicator, OGB-1. As expected, we see a wave of calcium fluorescence that extends deep into the dendrites. The amplitudes of the bAP-associated calcium transients exhibit several behaviors which reflect its dependence on the bAP. The calcium transients decrease in amplitude with distance from the soma, and on the dendrites, they also show a decrease in amplitude with successive spikes in a train. Furthermore, it is responsive to the transient K^+ channel blocker, Ba^{2+} . We also see amplification of the bAP evoked calcium signal, which is restricted to the distal dendrites, in response to pairing with presynaptic stimulation. This is consistent with extending the distance of propagation of the bAP.

To analyze these data, I have developed two new strategies for time series analysis of the fluorescent response to a train of action potentials. Furthermore, I have

implemented a method for functional data analysis that works for data defined on a branched domain. The end result of this analysis is a smooth functional representation of the degree of activation for all parts of the cell within the field of view of the microscope.

In another experiment, we examine the abstract neuron model called the clusteron, which learns to recognize specific input patterns by structural rearrangement of its synapses. We demonstrate that the model can also learn sequences of input patterns, and perform an exclusive-or operation: two tasks not typically solvable by single neuron models.

Contents

1	Introduction	1
2	Learning by spatial reorganization in an abstract neuron model	8
2.1	Introduction	8
2.2	Methods	9
2.2.1	Structure of the clusteron	10
2.2.2	Learning	13
2.2.3	Sequence presentation	15
2.3	Results	16
2.3.1	Distribution of synaptic activations and learning	17
2.3.2	Sequence learning	21
2.3.3	Feature detection	24
2.3.4	A reduced model	27
2.4	Discussion	30
3	Patterns of back-propagation in live neurons	37
3.1	Methods	41
3.1.1	Electrophysiology & calcium measurement	41
3.1.2	Data acquisition and analysis	43
3.1.3	2D and 3D scanning	43
3.1.4	Experimental protocol	47
3.1.5	Calcium transient extraction	47
3.1.6	Spatial functional data analysis	50
3.2	Results	57
3.2.1	Three-dimensional imaging of bAPs	58
3.2.2	Linearity of the fluorescence response	59
3.2.3	Pointwise analysis of fluorescence signals	61
3.2.4	Functional data analysis on the apical trunk	63
3.2.5	Functional data analysis on the entire tree	66
3.3	Discussion	68
4	Data analysis by sequential Monte Carlo particle filter	72
4.1	The model of the fluorescence signal	75
4.1.1	Fluorescence as a function of $[Ca^{2+}]_i$	76

4.1.2	Variability in PMT output	78
4.1.3	PMT voltage as a function of fluorescence	80
4.2	Particle filter estimation of the calcium signal	86
4.2.1	The SMC particle filter algorithm	87
4.2.2	Summary of parameters	90
4.3	Performance with synthetic data	91
4.3.1	Estimating the amplitudes of the transients	94
4.3.2	Generating a filtered trace	96
4.3.3	Comparison to low-pass filtering	97
4.4	Conclusions	99
5	Discussion	101
5.1	Summary	101
5.2	Coincidence detection	102
5.3	Cluster sensitivity	103
5.4	Model-based analysis	105
	Bibliography	108

List of Figures

2.1	Schematic of the clusteron	12
2.2	Distribution of synaptic activations	18
2.3	Dynamics of single pattern learning	20
2.4	Learning a sequence of patterns	23
2.5	Feature detection	26
2.6	Performance of the reduced model	31
3.1	Schematics of the optical setup	44
3.2	Schematics of the experimental setup	48
3.3	Description of the GLM	49
3.4	Construction of a spline function	51
3.5	Synthetic data and the basics of functional data analysis	55
3.6	Simple branched spline basis functions	56
3.7	3D scanning live neurons	58
3.8	K-channel dependence of bAPs	60
3.9	Single point examples of fluorescence traces	62
3.10	Single spike propagation and boosting	64
3.11	Spike train propagation through the apical trunk	65
3.12	Branched functional data analysis	67
4.1	Problems with traditional analysis techniques	74
4.2	Relationship between membrane potential and fluorescence	76
4.3	Schematic of a PMT	80
4.4	Noise profile of PMT output	81
4.5	Construction of, and inference from, the noise model	83
4.6	Generation of a synthetic dataset	93
4.7	Estimating the amplitude by SMC particle filter	95
4.8	Estimating $[F]_{total}$ by SMC particle filter	95
4.9	Comparison to a digital low-pass filter	98

List of Tables

4.1	Fixed model parameters	90
4.2	Monte Carlo estimated parameters	91

Chapter 1

Introduction

Neuroscience embodies many of the great mysteries that are still unsolved by modern science. At very small temporal and spatial scales, the human brain is governed by chemical reactions. Ion movement is controlled by local electrical and concentration gradients, which drive them in one direction or another. Lipid molecules align their non-polar tails and form bi-layer membranes. This happens because of the favorable energetic relationship of avoiding the polar aqueous solvent. Molecules form ionic bonds with other molecules, based on electrostatic interactions.

Altogether, the adult brain consists of approximately three pounds of water, lipids, proteins, salts, and sugars, etc., interacting in countless chemical reactions of this sort. The result is an organ that integrates information from at least five major senses, to generate a three-dimensional, object-oriented representation of the world. Another emergent property of this three pound lump is an ability to recall previous events and to predict possible future events. It enables internal visualization and dialog, concept abstraction, setting goals, and long term planning.

Much is known about the first stages of information consolidation and processing.

For example, the signal transduction pathways that enable sensory input to enter the brain are very well studied and understood. In the retina, light energy is converted to a change in the concentration of cyclic GMP and subsequently a change in membrane voltage of the photoreceptor. Photoreceptors drive bipolar cells and horizontal cells, which in turn stimulate amacrine cells and retinal ganglion cells. The ganglion cells then send the information out of the retina and into the central nervous system. (For basic neuroscience, See Kandel et al. [44] or Squire et al. [76])

Sensory input at the early stages is closely related to the nature of the organ which detected it. Neurons in the early stages of visual cortex are highly sensitive to light input in one specific region of the visual field, but in many cases, they have already developed a preference for bars of a specific orientation. Higher levels of visual system processing yield neurons which are most sensitive to more complex features, such as faces or objects. Amazingly, these neurons have even lost their spatial sensitivity, insofar as they can respond to a specific object, regardless of the size of its projection on the retina, or the location within the visual field.

As in the example above, the behavior of many types of neurons has been studied and described in detail. What lacks in our understanding is how the neurons have attained their specific firing preferences, and what features of the input are most important. Specifically, one area where knowledge of the system is lacking is in the connectivity between individual neurons, and the subcellular locations of the connections. Learning at the level of single neurons may indeed reflect changes in both the strength and the spatial pattern of synaptic connections (For a review see [8]).

Recent experimental and theoretical studies have helped establish the importance of active dendritic properties in information processing (For reviews see [24, 34, 53]). It is now well known that dendrites contain a host of voltage-dependent conductances (see [30, 36, 39, 55, 91, 93] and [43] for review), which play an important role in coincidence detection [79] and normalization of temporal summation [55]. Moreover, theoretical work has suggested that nonlinear summation of inputs greatly increases the memory capacity of neurons [62, 63].

Nonlinear summation of synaptic conductances by active dendrites, along with the decay of synaptic potentials with distance from the site of transmitter release, imparts the cell with a sensitivity to the distribution of their inputs [49, 54, 60, 64, 90]. Spatial patterns of synaptic inputs containing clusters of nearby synapses can activate voltage-dependent currents more strongly than patterns with distributed synapses. Thus, a rearrangement of synaptic positions along the dendritic tree can profoundly alter the response of the postsynaptic neuron even when other characteristics of the presynaptic input are unaltered [60, 63].

Despite these observations, systems that learn by synaptic rearrangement have received little attention. An exception is the clusteron [60]. The clusteron abandons all changes in synaptic weights in favor of learning by synaptic rearrangement. It is thus an excellent choice to evaluate the strengths and limitations of synaptic reorganization as a learning paradigm. Chapter 2 presents our theoretical work on this abstract neuron model.

As introduced above, we study neurons primarily from the point of view of a

single cell. We ask such questions as: How does a neuron decide when to fire? What patterns of input and output are ideal for inducing a change in the strength of specific inputs?

Answering these questions is problematic in several ways, due to the small size of the structures involved, and the rapid timecourse of the signals in question. Neurons, especially the large pyramidal neurons of the hippocampus and neocortex, are not isopotential [65]. Since the membrane potential depends on the location along the dendrite, we must have the capacity to record the activity level at different points of those structures. This can be problematic, since dendritic branches are often less than $1\mu\text{m}$ in diameter. Furthermore, action potentials, and post synaptic potentials are fast (an action potential occurs in about 2ms). The small structures and rapid signals necessitate advanced instrumentation to record such events.

Some basic ideas have emerged from the large amount of research at this level of questioning. For instance, Hebb's rule, often paraphrased as "cells that fire together, wire together" has been very influential since it was first introduced in 1949 [35]. This idea, that the connection between neurons that are coincidentally active will tend to be strengthened over time has been well justified in many cell types, although some cells follow different rules [61]. While very simple, this rule can be shown to result in several known features of neural connectivity, such as cortical map formation [7]. At a single cell level, Hebb's rule may be defined in an even more concrete way. That is to say, given a pair of cells, one presynaptic and one postsynaptic, the connection between the two has been found to increase in strength when the presynaptic cell fires

a short time before the postsynaptic cell [15]. In contrast, the connection between two cells decreases in strength when the opposite pattern of firing occurs, *i.e.* post before pre. This type of plasticity is referred to as spike-timing-dependent plasticity, STDP.

Action potentials are usually thought to initiate at the soma, or nearby on the proximal part of the axon, and then propagate down the axon to stimulate other cells. However, it is known that the action potential can and will propagate “backwards” through the cell, down the cell’s own dendrites [30, 80, 81]. This back-propagating action potential, bAP, may be the messenger that conveys information about the output firing of the cell back to the synapses [79]. In this model of neural behavior, it is on the dendrites where there is a convergence of information regarding the activity level of both pre and postsynaptic cells. Both requirements of coincidence detection are locally available, and a decision regarding whether the synapses should be strengthened or weakened can be made.

Several key pieces of information are known about the behavior of bAPs. First, bAP propagation is different from axonal propagation, since the amplitude and shape of the bAP change with distance. More specifically, bAPs propagate decrementally, getting smaller (and wider) with distance from the soma. This has been attributed to a change in the concentration and kinetics of a transient potassium current (the A-current or A-type potassium current) that gets larger and faster with distance from the soma [36].

Some have proposed, and there is some evidence, that the bAP can be amplified

on the dendrites by appropriately timed presynaptic stimulation [29, 79, 89]. This is thought to be mediated by the activation, and subsequent inactivation, of the A-current. Inactivation of the A-current changes the balance of excitatory and inhibitory ion channels, thereby making it easier for a bAP to propagate. This is a potential mechanism for coincidence detection, whereby only the nearly simultaneous activation of presynaptic fibers, and the postsynaptic cell will result in a large amplitude bAP. Furthermore, the guidance of action potential propagation may even be branch specific.

Another interesting phenomenon has been observed in measuring the response of the dendrites to a train of action potentials. It is known that action potentials recorded from the soma or axon are very reliable and their shape is nearly immutable. On the dendrites, however, bAPs will dramatically decrease with successive spikes [11]. That is, the second bAP will be smaller than the first, the third smaller than the second, and so on. This has been attributed to a difference in the behavior of the voltage-gated sodium channels between those on the axon and those on the dendrites. More specifically, while dendritic sodium channels activate and inactivate in a similar fashion to somatic sodium channels, they experience a slower recovery from the inactivated state. This results in fewer and fewer sodium channels being available for subsequent spikes in the train [11].

Some of the main questions addressed here in Chapters 3 and 4 are: What is the spatial pattern of back-propagation, and how does it change with presynaptic stimulation? How does the pattern of back-propagation change with spikes in a

train, and how does it change the region of boosting when paired with presynaptic input? If the region of boosting is different for different spikes in a train, the result could be a fairly complicated learning rule that depends on the spiking history of the cell, and the location of the synapses in question. This could change the rules for spike-timing-dependent plasticity for each spike in a train.

Chapter 2 of this dissertation involves the theoretical analysis of the capabilities of the clusteron model. Through the use of numerical simulation we describe two interesting tasks that the clusteron can learn, which are learning of a sequence of input patterns, and feature detection. Chapter 3 describes our experiments on the back-propagating spikes of hippocampal pyramidal neurons. This includes a brief description of the instrument that was developed for the task, and some novel analysis techniques developed to better make use of the exogenous information we have about the data collection system. The result of our experimentation and analysis confirms several findings about the behavior of bAPs, and extends our knowledge of how bAPs respond to patterns of presynaptic and postsynaptic stimulation. Chapter 4 extends the idea of maximizing our data analysis by including knowledge of the system, through the use of a full model description of how our signal of interest is manipulated by our experimental procedure, and corrupted by the noise imparted by our equipment. The model is then used by a sequential Monte Carlo particle filter for extracting the most likely signal, given the recorded output.

Chapter 2

Learning by spatial reorganization in an abstract neuron model

2.1 Introduction

Neurons receive and integrate information in the form of synaptic conductances across their dendritic trees. Synaptic input is characterized by the spatial and temporal distribution of active synapses, and by the strength and timecourse of individual synaptic conductances.

The goal of the present study was to develop a mathematical framework that can be used to describe and analyze the clusteron and other models that learn through structural changes. Within this framework we investigated the mechanisms and learning rate characteristics of the clusteron in two basic configurations: one similar to the original clusteron [60], and one consisting of discrete bins of integration, abstractions of the computational subunits of a dendritic tree [63, 64, 90].

Once the mechanism of learning is understood, the model can be extended to perform other tasks. We first show how the introduction of simple temporal dynamics results in a model capable of learning spatio-temporal patterns. We then show how the

conditions for synaptic rearrangement can be changed to create models that either respond to a common feature in the set of training patterns, or the distinguishing features of each pattern in the set.

Our analysis reveals a general feature of synaptogenesis: the length of time needed to learn an association, and the ultimate strength of the association, are both dependent on the likelihood of forming an appropriate synaptic connection. The contributions of the individual synapses to the output follows a Gaussian distribution, and an approximation can be made that displays the relations to learning rate and association strength explicitly. Analyzing the distribution of synaptic activations, yields a clearer understanding of the principles underlying learning by structural rearrangement.

2.2 Methods

The clusteron differs from most other neuronal models in the literature in that learning is a result of changes in the physical arrangement of synapses on the cell, rather than changes in the individual synaptic weights. Its primary feature is that the contribution of an individual synapse to postsynaptic activation (e.g., depolarization) is modulated by a nonlinear function of the total number of active synapses in its vicinity. Because the number of active synapses and the synaptic weights are kept constant, only changes in the spatial pattern of synaptic activation result in differential postsynaptic responses.

2.2.1 Structure of the clusteron

We consider a dendritic tree on which N synaptic connections have been formed by N numbered, afferent fibers. We refer to a collection of fibers that are activated by a stimulus as an *input pattern*. Mathematically, an input pattern $\mathbf{v} \in \{0, 1\}^N$ is an N -vector of 1s and 0s so that if $v_i = 1$, then the i -th afferent fiber is active in the given pattern. The synaptic connections these fibers make on the cell are described by an injective function $\phi : \{1, 2, \dots, N\} \rightarrow \{1, 2, \dots, N\}$ that is independent of the input pattern \mathbf{v} . Therefore, $\phi(i) = j$ implies that fiber i innervates the synapse j , and, similarly, $\phi^{-1}(j) = i$ implies that synapse j receives an input from fiber i . During training, the set of input patterns $\{\mathbf{v}^i\}$ remains unchanged, meaning that each pattern \mathbf{v}^i represents firing of a consistent population of presynaptic cells, *i.e.* activation of a consistent set of fibers. However, the locations at which afferent fibers innervated the cell, represented by ϕ , did change during learning. For notational reasons it was easier to change the inverse ϕ^{-1} rather than ϕ directly.

We used two versions of the model: The first, illustrated schematically in Figure 2.1(b), was similar to the original clusteron [60] and contained a cell body and a single dendrite with many synapses. The impact of an active synapse was boosted as a function of the activity of other synapses that lay within a certain radius. The second, “branched” model, presented schematically in Figure 2.1(c), can be thought of as a reduction of a two-layer model of dendritic integration [62]. Rather than a single dendrite, it contained non-overlapping regions that partitioned the dendrite into “bins”,

which can be thought of as separate dendritic branches [63, 64]. Synapses within a branch interacted nonlinearly, but the total depolarization was a linear combination of individual branch activity. The two variants behaved similarly, but differed in several important ways discussed below.

The total somatic activation, analogous to the membrane potential at the cell body induced by an input pattern, was calculated as follows: In the first model the activity of synapse i was modulated by all synapses within a given physical radius of its location. We denoted the set of all synapses that affected the activity of synapse i by D_i . Since synapses were numbered sequentially, we set $D_i = \{1 \leq j \leq N | i - K \leq j \leq i + K\}$ where K is the radius of D_i (see Figure 2.1(b)). The activation due to synapse i , given an input pattern \mathbf{v} , was

$$a_i(\mathbf{v}, \phi) = v_{\phi^{-1}(i)} \sum_{j \in D_i} v_{\phi^{-1}(j)}. \quad (2.1)$$

Therefore, the synaptic activation was the product of synapse i 's own input, either 1 or 0, and the sum of all other inputs in D_i . This can be generalized to $a_i(\mathbf{v}, \phi) = v_{\phi^{-1}(i)} F\left(\sum_{j \in D_i} v_{\phi^{-1}(j)}\right)$. The form of F modulates the summation, and it can be chosen to model sublinear spatial summation, as in the case of a passive cell [65]. We only considered $F(x) = x$ and typically chose 1000 input lines and a radius of integration that included 20 synapses or more.

In the branched model, the activation of whole branches, rather than single synapses is used to determine the somatic activation. In particular, the activation of

a.

$$\mathbf{V} = \{0, 0, 1, 1, 0, 1, 0, 0, 1\}$$

$$\begin{array}{c} \phi \\ \downarrow \\ \{1, 0, 1, 0, 0, 1, 1, 0, 0\} \\ \uparrow \\ \phi^{-1} \end{array}$$

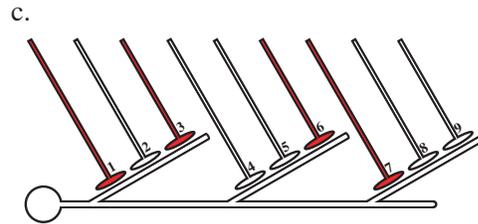
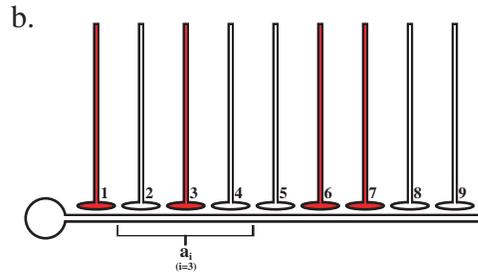


Figure 2.1: **Schematic of the clusteron.** (a) The function ϕ , maps the input vector, \mathbf{v} , to locations on the dendrite. The training paradigm modifies ϕ . In this example: $\phi(1)=5$, $\phi(2)=2$, $\phi(3)=6$, $\phi(4)=3$, $\phi(5)=8$, $\phi(6)=1$, $\phi(7)=9$, $\phi(8)=4$, $\phi(9)=7$. (b) Schematic of the original clusteron with a single dendrite. Active inputs are shown in red and are consistent with the example mapping shown in panel (a). Total synaptic activation is determined by the activity of nearby synapses (see Equation 2.1). Shown is window D_3 of radius $K = 1$ around synapse 3. (c) Schematic of the branched version of the clusteron. The input vector is mapped onto the ‘branches’ of the cell. Nonlinear interactions occur only within a branch (see Equation 2.2), so that synapses 6 and 7, which would interact in the unbranched case, now do not interact. Conversely, synapses 1 and 3 now do interact, whereas in the unbranched case they were too far apart.

branch m was given by

$$b_m(\mathbf{v}, \phi) = \left(\sum_{j=1}^k v_{\phi^{-1}(j)} \right)^2. \quad (2.2)$$

As in the previous case, this can be generalized to $b_m(\mathbf{v}, \phi) = G(\sum_{j=1}^k v_{\phi^{-1}(j)})$. We chose $G(x) = x^2$ for consistency with the unbranched version of the model.

The total somatic activation was obtained as a sum of all N individual synaptic activations in the first, and as a sum over all M branch activations in the branched model. In particular, the depolarization $W(\mathbf{v}, \phi)$ at the soma due to an input pattern \mathbf{v} and an arrangement of afferent fibers ϕ was given for the two models respectively by

$$W_1(\mathbf{v}, \phi) = \sum_{i=1}^N a_i(\mathbf{v}, \phi), \quad \text{and} \quad W_2(\mathbf{v}, \phi) = \sum_{i=1}^M b_i(\mathbf{v}, \phi). \quad (2.3)$$

2.2.2 Learning

In both models, a supervised learning protocol selectively stabilized the most highly active synapses. The training protocol was divided into a number of “epochs”, each consisting of alternating presentation of the training patterns followed by a judgment of synaptic suitability. An epoch ended with the spatial rearrangement of poorly performing synapses.

More precisely, let a_i^j be the activation of synapse i in response to the j -th pattern in an epoch. In the first model the average activation \bar{a}_i over all P training patterns presented during an entire epoch, $\bar{a}_i = 1/P \sum_{j=1}^P a_i^j$, was compared to a threshold value ζ . If $\bar{a}_i > \zeta$, the fiber afferent to the synapse was fixed, *i.e.* $\phi^{-1}(i)$ remained unchanged. For a synapse, ζ represents the level of activity that it must achieve

in order to become stabilized, analogous to the degree of local activation needed to induce long term potentiation. In these simulations, the choice of ζ dictates the probability that a synapse can become stable.

The indices i of all synapses whose activation was insufficient, that is $\bar{a}_i \leq \zeta$, formed a set R . To redefine ϕ^{-1} on R , we chose a bijective function $H : R \rightarrow R$, and redefined ϕ using

$$\phi_{new}^{-1}(i) = \begin{cases} \phi_{old}^{-1}(i) & \text{if } \bar{a}_i > \zeta \\ H(i) & \text{if } \bar{a}_i \leq \zeta. \end{cases} \quad (2.4)$$

We chose H randomly, and allowed it to change between training epochs [60]. Different choices for H reflecting the targeting of certain locations on the dendrite could also be considered [31].

Synapses located in regions with a higher density of active synapses, i.e. active clusters, would attain higher values of synaptic activation and were rewarded by stabilization. Isolated synapses had smaller activations, and were moved to potentially join established clusters or nucleate new ones. This protocol was iterated throughout the simulation.

The protocol for the branched model was similar; however suitability was determined at the branch level: If the average branch activation over a training epoch (see Equation 2.2) exceeded a threshold, then all synapses on that branch were stabilized. If not, all synapses on the branch became part of the pool R , and were reshuffled according to Equation 2.4. Consequently, synapses were only stabilized by collectively pushing a branch activation over threshold, and not by joining an existing

stable branch. This assumption simplified the subsequent analysis. Alternate forms of learning in the branched model resulted in qualitatively similar behavior.

The choice of the threshold ζ had a large impact on model performance. We will discuss cases in which ζ was fixed, and cases in which ζ varied as a function of the average synaptic activation. In the second case, a new form of synaptic competition is introduced: The stabilization of some synapses may cause an increase in the learning threshold and lead to the destabilization of previously stable synapses. These newly destabilized synapses then have to compete for position on the dendritic tree that will again lead to stabilization. The value of the variable threshold was typically given by the mean synaptic activation, or a fraction thereof.

2.2.3 Sequence presentation

Both the standard and branched models could be extended to allow for presentations of spatio-temporal patterns. Sequences $\mathbf{V} = (\mathbf{v}^1, \dots, \mathbf{v}^L)$, of spatial patterns \mathbf{v}^i described above, were presented during each training epoch. The somatic activation $W(\mathbf{p}^n, \phi)$ in response to the n -th pattern in a sequence was obtained using Equation 2.3 and

$$\mathbf{p}^n = \mathbf{v}^n + \alpha \mathbf{p}^{n-1} \quad 0 < \alpha < 1, \quad (2.5)$$

and $\mathbf{p}^0 = \mathbf{v}^0$. A fraction α of the raw input due to the preceding pattern was held over to compute the synaptic activation of the present pattern. Thus, the activation function reflected not only the spatial contiguity, but also temporal contiguity of synaptic activations. We considered a spatio-temporal pattern to have been learned

if the somatic activation of the training sequence exceeded the activation induced by all other permutations of the training sequence, as well as those of a sequence of random patterns.

The learning algorithm was equivalent to that described in the previous section. An epoch consisted of a single presentation of the training sequence. However, the synaptic activation upon the presentation of the *final* pattern in the sequence, rather than average activation, was used to determine the suitability of a synapse. Therefore, if the training sequence consisted of L patterns, then all synapses satisfying $a_i^L < \zeta$ were reshuffled.

2.3 Results

We next present an intuitive description of learning in the two versions of the clusteron, and use these insights to develop and analyze several extensions of the learning rule. As implied by the name, the spatial clustering of synapses was crucial for the correct recognition of learned patterns [60]. The nonlinear interaction between clustered synapses resulted in higher somatic activations than those evoked by arbitrary patterns (Figure 2.2(a)). We make these observations more precise by considering the distribution of the activations of synapses in the model. The evolution of this distribution during training is then fully described in a reduced model.

2.3.1 Distribution of synaptic activations and learning

We begin by describing the distribution of synaptic activations in the clusteron and how it changes during training. Note that the activation of synapse i , given by Equation 2.1, is directly proportional to the number of active synapses in its neighborhood. Therefore, the degree of clustering of synapses activated by an input pattern can be represented by the frequency histogram of synaptic activations.

Since, in our simulations, the total fraction of active synapses was small (typically 15%), and the windows of interacting synapses were large (typically $K = 20$), synaptic activations approximately followed a binomial distribution, parameterized by the number of active synapses N_{active} , and the probability of randomly choosing a specific window of integration (*i.e.* $\frac{K}{N}$). Since the number of active synapses is large, this binomial distribution was well approximated by a normal distribution. In untrained patterns, the synapses are distributed randomly in space, which results in an approximately normal distribution of activations (Figure 2.2(b)).

Note, synapses not active in any patterns are not affected by restructuring (see Equation 2.1). The large peak at 0 due to such synapses was omitted from synaptic activation histograms for clarity.

Histograms of synaptic activation (e.g., Figure 2.2(b)) demonstrated several important features. Patterns that activated clusters, contained a higher number of highly activated synapses. The corresponding distributions therefore lie to the right of those corresponding to patterns activating a random subset of synapses (Figure 2.2(b)).

The learning threshold, represented by a vertical line in all figures, separated synapses to the left that were reshuffled, and those to the right that are fixed at the end of a training epoch (see Equation 2.4). Figure 2.2(b) shows the result of training in the case of a fixed learning threshold. As synapses were reshuffled randomly, they occasionally experienced increased activation due to joining an existing cluster or nucleating a new one. If this activation was above the learning threshold, the synapse was fixed. Therefore, during training the learning protocol resulted in a gradual rightward shift of the distribution of synaptic activations. When the activation of all synapses lay above threshold the system reached equilibrium. The

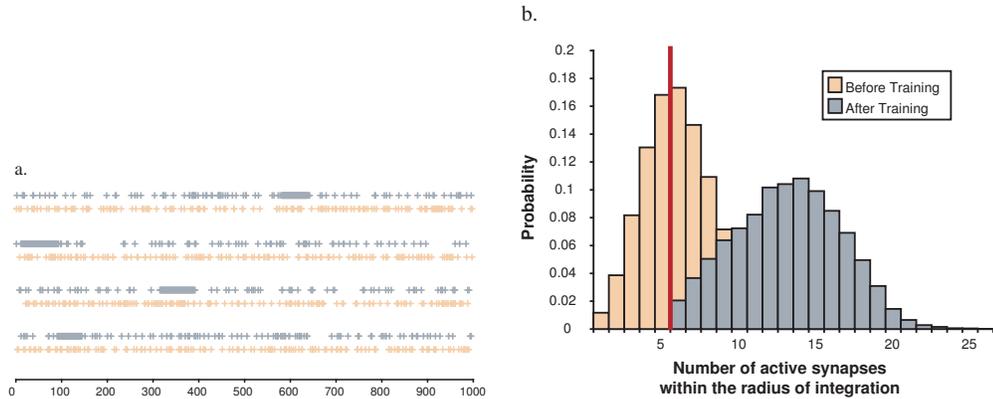


Figure 2.2: **Distribution of synaptic activations.** (a) The location of the 150 active synapses activated by each of four training patterns before (lower of the pairs) and after (higher of the pairs) training. After 100 training epochs, the active synapses formed clusters along the dendrite. The x -axis represents distance along the dendrite. (b) Histograms of synaptic activations for all active synapses before and after training. Before training, the activations approximately followed a normal distribution. The vertical line represents the learning threshold used in this simulation. During training, synapses formed clusters, causing a rightward shift in the distribution. Note, the peak of synaptic activations corresponding to 0 does not change with training and was removed for clarity.

choice of the learning threshold was critical in determining how well and how fast the model learned. High thresholds lead to the best learning, *i.e.* the largest increase in somatic responses to the training patterns. However, the time for an equilibrium to be reached was typically long. Alternatively, low thresholds lead to rapid learning, but resulted in relatively small increases in the somatic response. Figure 2.3(a) shows examples of the somatic response during the course of training to a single pattern, under three different threshold levels. The branched clusteron and other variations that were tested show the same relationships between learning speed and magnitude to the learning threshold (See Figure 2.3(b)).

This relationship can be explained intuitively by considering the effect of the threshold on the evolution of the synaptic activation distribution. If the threshold is low, the main mass of the distribution would lie above it. Moreover, even synapses with activations below threshold would only need a small boost to cross it. Therefore, training resulted in rapid equilibration, but only an incremental increase in the total activity. In the case of a high threshold synapses needed a large boost in their activation to be stabilized. Random reshuffling rarely resulted in such large increases, and equilibrium was reached more slowly. However, once the distribution lay to the right of a high threshold the total activation could be very high.

This argument assumes that the learning threshold is constant during training. However, as discussed in Section 2.2.2, the threshold can be set to increase with the magnitude of the somatic activation, in a way similar to the BCM learning rule [5]. Such increasing thresholds leading to a higher degree of competition between synapses:

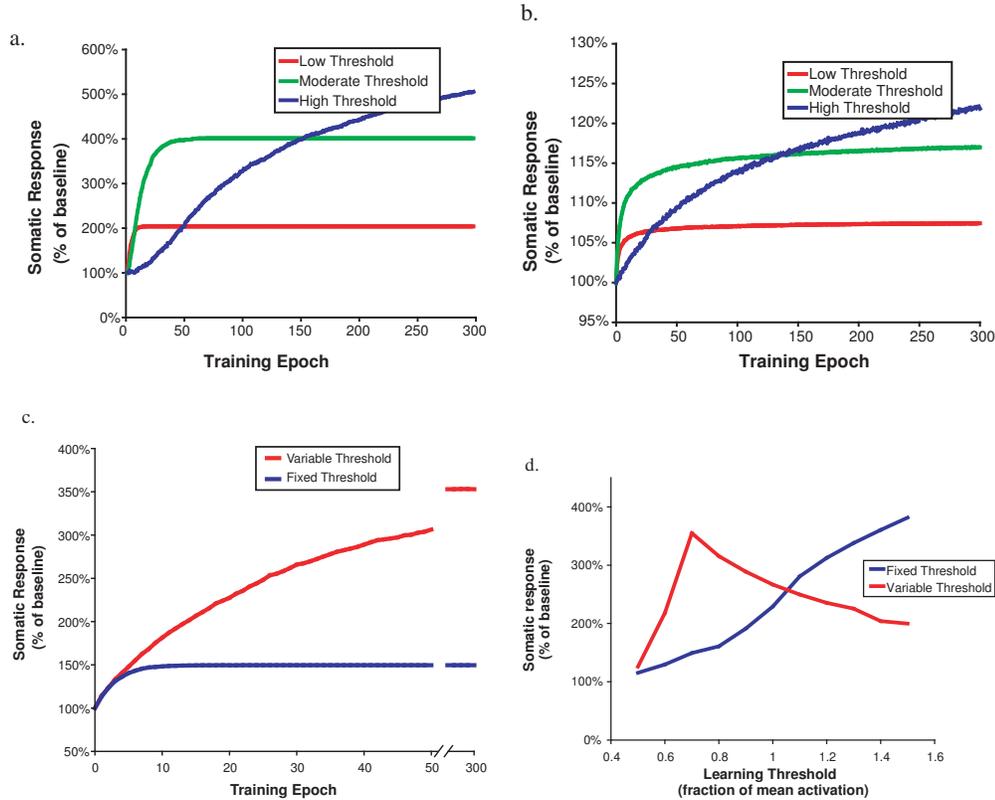


Figure 2.3: **Dynamics of single pattern learning.** (a) The increase in the somatic response to training on a single pattern strongly depended on the learning threshold. Low thresholds lead to rapid learning, while high thresholds resulted in longer equilibration times. However, high thresholds resulted in the strongest responses to learned patterns. (b) The branched model showed the same relationships for learning speed and magnitude. The vertical scale is different in the two cases. Since all synapses on a branch are fixed after its activation exceeds threshold, the increase in the activation during training is smaller in the branched model. (c) A strong response could be evoked rapidly by using a low variable threshold. Here, the initial learning thresholds are equal, but the variable threshold rises as a function of the mean synaptic activation, $\zeta = 0.7\bar{a}_i$. Note how learning is of a comparable rate for both cases, yet the variable threshold ultimately allows for much better learning. (d) Variable thresholds grant an advantage in learning when they are relatively low. If the variable threshold is too high, very few synapses are ultimately allowed to be stable, which limits the total somatic activation that can be achieved. On the other hand, the steady-state degree of learning is monotonically related to the learning threshold in the fixed case, although the rate of learning suffers, as described in this manuscript.

If the threshold increased sufficiently rapidly, at the end of each training epoch only a fixed fraction of the synapses was stabilized. Therefore, the activity of synapses that were stable at the end of a previous epoch could fall below the increasing threshold as they are outperformed by newly reshuffled synapses. In contrast, with a fixed threshold, all synapses can be stabilized at the same time, provided the learning threshold is not prohibitively high. More precisely, if $\zeta \leq K$, the radius of integration, there will exist at least one absorbing state in which all synapses will be stabilized.

Such variable thresholds can result in rapid learning and a large response to learned patterns (Figure 2.3(c)). During the early stages of training, the threshold was low, allowing rapid nucleation of clusters. During the later stages the threshold increased with the mean synaptic activation, and resulted in large increases in the learned response. This relationship holds for a range of values for the variable threshold, but fails when the threshold is very high (Figure 2.3(d)). In that case, only very few synapses are allowed to become stabilized and the total somatic activation that can be achieved suffers.

2.3.2 Sequence learning

The observations made in the previous section can be used to obtain a modification of the clusteron algorithm capable of learning spatio-temporal patterns of input. Upon following the learning paradigm described in Section 2.2.3 with a sequence of inputs $\mathbf{V} = (\mathbf{v}^0, \mathbf{v}^1, \dots, \mathbf{v}^L)$, the presentation of the sequence of patterns in correct order resulted in the largest somatic activation (see Figure 2.4).

The distribution of synaptic activations again clarifies the underlying mechanism. Equation 2.5 indicates that the contribution of a pattern to the activation of a synapse decreases exponentially in time. If each pattern activated a small subset of synapses, then after the presentation of the n -th pattern the distribution of synaptic activations could be decomposed into n parts, each corresponding to one pattern in the sequence (see Figure 2.4(b)). When latter patterns in the sequence were presented, synapses activated by earlier patterns had decayed to the lower part of this distribution. Since the activation of each synapse was compared to a single learning threshold at the time of presentation of the last pattern, each pattern had undergone a variable degree of exponential decay, and therefore experienced a different drive to cluster. As a result, synapses activated by latter patterns in the sequence became the most clustered and resulted in the largest single pattern responses. Thus, the training sequence was then represented on the cell model by patterns of increasing clustering, which resulted in the largest somatic responses (see Figure 2.4(c)).

Interestingly, changes in the choice of threshold had a large impact on this outcome. For example, a fixed threshold resulted in higher responses to the sequence \mathbf{V} presented in reverse order, while a variable threshold, resulted in a preference for the proper order. The explanation of this is that with a fixed threshold, the steady-state degree of clustering was similar to the degree of clustering when the model was trained with a single pattern, resulting in the best clustering of patterns with the highest relative learning threshold. On the other hand, when only a limited number of synapses can be stabilized at a time, an advantage was gained by latter patterns

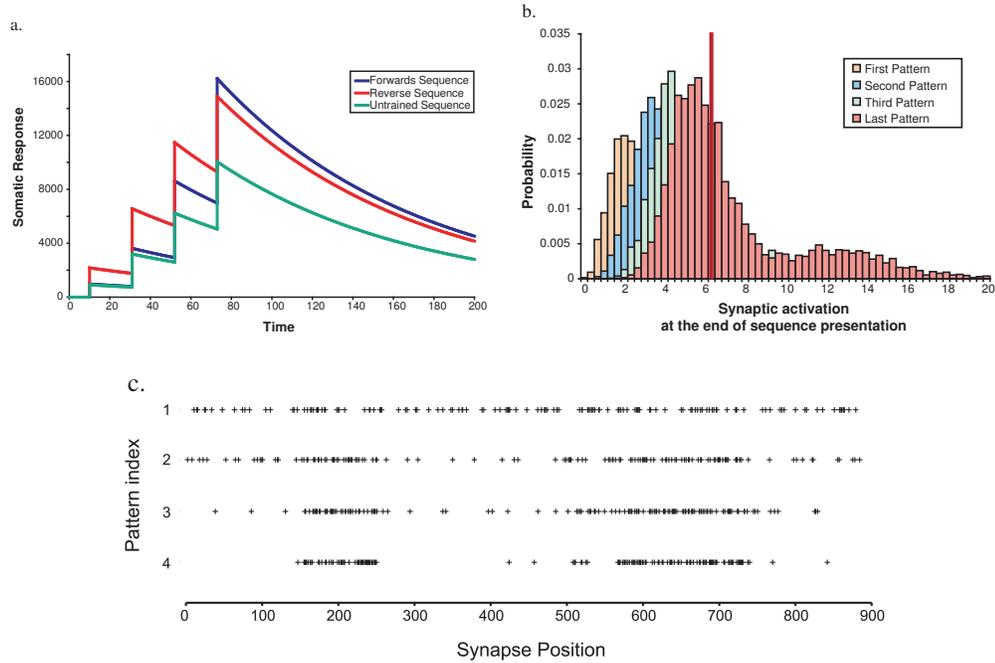


Figure 2.4: **Learning a sequence of patterns.** (a) The response of the clusteron to a learned spatio-temporal sequence. The response to the sequence presented in the proper order was highest. Also shown is the response to the sequence presented in reverse order (red trace) and the response to a sequence of four random patterns. (b) The distribution of synaptic activations right after the presentation of the last pattern is multi modal. Synapses activated by earlier patterns in the sequence have activation that lie farther below the learning threshold (the vertical bar) and experienced a smaller drive to cluster. (c) The location of active synapses for the four patterns in the training sequence after training. Note that clusters for different spatial patterns formed on overlapping regions of the dendrite, so that synaptic activation was boosted when the patterns were presented sequentially.

in the sequence, that clustered rapidly because of a low relative learning threshold. In both cases, the preferred sequence was one in which patterns are presented in the order of increasing clustering.

2.3.3 Feature detection

A system trained to respond to a set of patterns, such as faces, can do so in two distinct ways: It can respond to a *specific feature* of each pattern in the training set (such as a scar or other distinguishing mark). Alternatively, the system can respond to a feature *shared* by all patterns in the training set (all faces in the training set may feature a nose). In this section we show that the clusteron can be trained to respond to either the shared features or specific features of the patterns in the training set.

As an example, consider the two patterns shown on the top of Figure 2.5(a). The two input patterns in the figure each activated 25% of the fibers that synapsed on the dendrite. Moreover, half of the fibers activated by one pattern are also activated by the other. The fibers activated *only* by pattern 1 are denoted s_1 , those activated only by pattern 2 are denoted s_3 , and those activated by both patterns are denoted s_2 . All of these sub-patterns consist of 12.5% of the total fibers. Also shown are pattern s_4 composed of fibers not activated by either pattern in the training set, and a random pattern of 12.5% of the fibers. Note that this figure illustrates the patterns of fibers that are activated, and does not indicate the location of the activated synapses on the dendrite. While a pattern always activates the same fibers, their synaptic contacts change during training.

Pattern s_2 can be obtained by performing a logical AND operation on the two training patterns, and represents their common or shared features. Similarly, patterns s_1 and s_3 can be obtained by obtaining a logical XOR operation and represent specific features of the first and second training pattern respectively. Figure 2.5(b) illustrates that different choice of learning threshold will lead the clusteron to preferentially respond to either shared or common features of the training set.

An examination of the distribution of synaptic activations again reveals the mechanism behind this type of learning. Figure 2.5(c) shows a histogram of synaptic activations used to decide which synapses are reshuffled at the end of the training epoch. Since the total synaptic activation is *averaged* over a training epoch (see Section 2.2.2), the distribution is bimodal. One part consists of synapses activated by only one of the patterns and contains $2/3$ of the total mass of the distribution. The other part consists of synapses activated by both patterns.

If the threshold is high, only synapses participating in both patterns were likely to attain average activations exceeding threshold upon reshuffling. Therefore, only synapses activated by the pattern s_2 were likely to be stabilized and experience a drive to cluster. Similarly, if the threshold was low, synapses participating in both patterns typically had average activations that already exceeded threshold. Therefore, only synapses activated by a single pattern experienced a drive to cluster that lead to an increase in activation.

We note that this effect depends crucially on the assumption that synaptic activation is averaged over an entire training epoch. Alternatively, we can normalize the

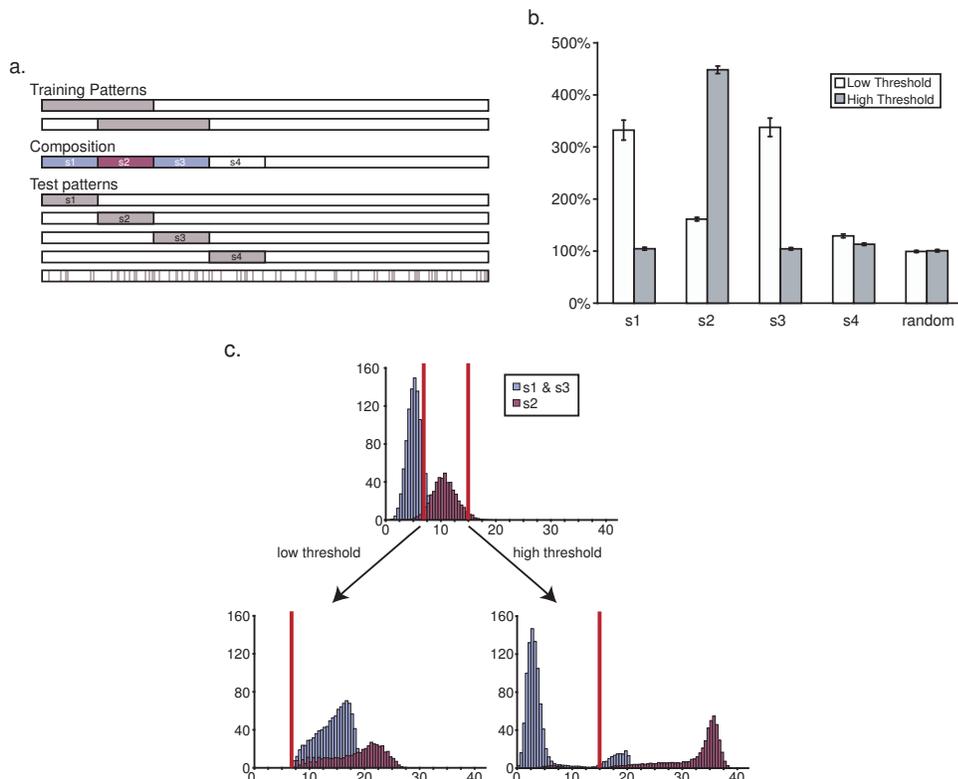


Figure 2.5: **Feature detection by the clusteron.** (a) Schematic of the training and testing patterns, showing the four possibilities for the activity of each synapse. Out of 4,000 fibers, 1,000 were activated by each training pattern, with 500 activated by both patterns. A fiber could be active by both training patterns ($s1$ and $s3$), one pattern only ($s2$), or neither pattern. The fifth test pattern is composed of 500 randomly chosen fibers. (b) Bar graph of the total somatic response to the five test patterns, normalized to the response to random untrained patterns. Response to the test patterns reflect the degree of clustering for each subpopulation of synapses in isolation. As shown, training with a high threshold resulted in high responses to features common to both patterns, while a low threshold resulted in larger responses to specific features. (c) The histograms of synaptic activations for the training patterns illustrate the underlying mechanism. Synapses activated by both patterns have higher average activations requiring a high learning threshold to stimulate significant clustering. Similarly, synapses active in only one pattern were best stimulated to cluster by a lower learning threshold. The distributions after training show modifications that are due to clustering of the subpopulation in question, as well as from clustering of the other subpopulation, since the training patterns contain both. The degree of clustering for subpopulations in isolation is shown in (b).

activity by the number of patterns in which a synapse participates

$$\bar{a}_i = \frac{\sum_{j=1}^P a_i^j}{\sum_{j=1}^P \phi^{-1}(i, j)}.$$

In this case, the distribution of activations becomes unimodal, and the model will tend to respond to any feature of the training patterns. These two measures of synaptic activation could represent different timecourses of input integration. Averaging synaptic activation over time requires a memory of previous pattern presentations and could be explained by a long-lived biochemical change due to the pattern presentation. Normalizing synaptic activation would only require that there be no such changes, or merely that pattern presentation is sufficiently slow enough to outlive such changes.

Interestingly, this type of feature detection was not seen in the branched clusteron. Intuitively, the activity of entire branches is too coarse a measure to discriminate patterns at the level of single fiber activity.

2.3.4 A reduced model

We next developed an analytically tractable reduction of the branched model that described how the distribution of synaptic activations evolved towards a steady state with repeated presentations of a training pattern. The reduced and branched model were qualitatively similar, and exhibited the same trends in the speed and magnitude of learning.

Given a branched clusteron with B branches and A active synapses randomly distributed across the branches, the binomial theorem can be used to approximate

the distribution of the number of synapses per branch and hence the distribution of branch activations: Let $n_0 = A$ and $m_0 = B$ be the initial number of active synapses and branches respectively. The training protocol called for the redistribution of the fibers lying on insufficiently active branches, that is those that contain less than A_ζ active synapses. These fibers were redistributed among the same set of branches. This procedure was repeated after every training epoch. In particular, after the k -th training epoch, there were m_k branches that contained less than A_ζ active synapses per branch and thus were insufficiently activated. Here A_ζ is the integer part of $\sqrt{\bar{\zeta}}$ (see Equation 2.2). At the end of a training epoch the n_k synapses residing on these branches were then redistributed randomly amongst the same set of insufficiently active branches.

For simplicity we consider learning with a single training pattern. Since the synapses were redistributed randomly, we can think of the n_k synapses as balls that are being distributed with equal probability among m_k bins. The number of balls per bin follows a binomial distribution, which can be approximated by a normal distribution of mean μ_k and variance σ_k^2 where

$$\mu_k = \frac{n_k}{m_k} \quad \text{and} \quad \sigma_k^2 = \frac{n_k}{m_k} \left(1 - \frac{1}{m_k}\right).$$

Thus, the distribution of the number of active synapses per branch is approximately $m_k \mathcal{N}(\mu_k, \sigma_k^2)(x)$, where we use $\mathcal{N}(\mu, \sigma^2)$ to denote a normal distribution with mean μ and variance σ^2 .

Using this expression we approximate the total number of branches that will be

insufficiently activated at the end of the next training epoch as

$$m_{k+1} = m_k \int_{-\infty}^{A_\zeta} \mathcal{N}(\mu_k, \sigma_k^2)(x) dx = \frac{m_k}{2} \left[1 + \operatorname{erf} \left(\frac{A_\zeta - \mu_k}{\sigma_k \sqrt{2}} \right) \right]. \quad (2.6)$$

Note that the mean number of synapses per branch, amongst the insufficiently activated branches, is given by

$$\bar{n}_k = \frac{\int_{-\infty}^{A_\zeta} x \mathcal{N}(\mu_k, \sigma_k^2)(x) dx}{\int_{-\infty}^{A_\zeta} \mathcal{N}(\mu_k, \sigma_k^2)(x) dx} = \frac{m_k}{m_{k+1}} \int_{-\infty}^{A_\zeta} x \mathcal{N}(\mu_k, \sigma_k^2)(x) dx.$$

Since these are distributed amongst the m_{k+1} branches, the product $\bar{n}_k m_{k+1}$ yields the total number of synapses, n_{k+1} , on the unstable branches, as

$$\begin{aligned} n_{k+1} &= m_k \int_{-\infty}^{A_\zeta} x \mathcal{N}(\mu_k, \sigma_k^2)(x) dx \\ &= \frac{m_k \mu_k}{2} \left[1 + \operatorname{erf} \left(\frac{A_\zeta - \mu_k}{\sigma_k \sqrt{2}} \right) \right] - m_k \sigma_k \sqrt{\frac{1}{2\pi}} e^{-\frac{(A_\zeta - \mu_k)^2}{2\sigma_k^2}} \\ &= \mu_k m_{k+1} - m_k \sigma_k \sqrt{\frac{1}{2\pi}} e^{-\frac{(A_\zeta - \mu_k)^2}{2\sigma_k^2}}. \end{aligned} \quad (2.7)$$

Note that Equations 2.6 and 2.7 are a dynamical system whose evolution models the change in the mean and variance of the distribution of activations.

The number of unstable branches and the number of unstable synapses on those branches, can be used to compute the mean and variance of the new distribution by again invoking the normal approximation to the binomial distribution. Therefore, the total distribution G after training epoch k can be calculated as a sum of normal distributions:

$$G_k = \begin{cases} m_k \mathcal{N}(\mu_k, \sigma_k^2)(x) & \text{for } x < A_\zeta \\ \sum_{i=0}^k m_i \mathcal{N}(\mu_i, \sigma_i^2)(x) & \text{for } x \geq A_\zeta. \end{cases}$$

Note that only branches above threshold are stabilized, and so the part of the distribution of synapses per branch above A_ζ includes all synapses stabilized in the past. The portion of the distribution corresponding to unstable branches only consist of those synapses assigned during the last round of training. The calculated distributions after different numbers of training epochs are shown in Figure 2.6(a). Figure 2.6(b) shows the averaged branch activation distribution for the analogous branched clusteron (125 branches, 40 synapses per branch, 50% active synapses), before and after 20 rounds of training.

Figure 2.6(c) illustrates that the reduced model displays the same relationships for the speed and magnitude of learning observed earlier. Here we computed the total somatic activation from their distribution by $act = \int_{-\infty}^{\infty} x^2 G_k$.

Furthermore, the steady state magnitude of learning is higher in models trained with a higher learning threshold. Since most synapses are ultimately stabilized above threshold, the steady state magnitude of learning is proportional to the number of branches that are initially below threshold (*i.e.* $\int_{-\infty}^{\zeta} \mathcal{N}(\mu, \sigma^2)$). This value evaluates to an error function that is also monotonically increasing.

2.4 Discussion

A hallmark of biologically plausible neural networks is that the rules governing the efficacy of synaptic connections depend only on locally available information. Learning rules for adjustment of synaptic weights based on the correlations of pre- and postsynaptic activity at a synapse have received considerable experimental and theoretical

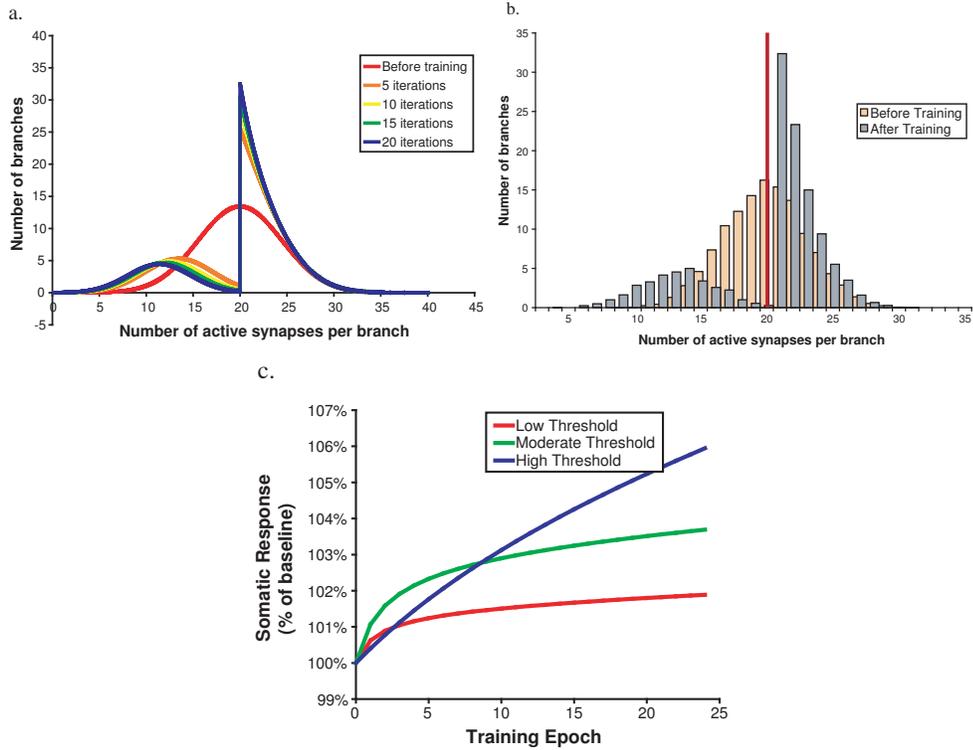


Figure 2.6: **Performance of the reduced model.** (a) The distribution of branch activations in the reduced clusteron model. The normal distribution at the onset of training as well as the distributions after a number of iterations of the reduced model are shown. (b) The distribution of branch activations of a similarly configured branched clusteron before, and after 20 rounds of training. Vertical red bar represents the learning threshold. (c) The distributions shown in (a) can be used to calculate the activation of the model during training. The performance of the reduced model demonstrates the same relationships for the speed and magnitude of learning dependent on the choice of learning threshold.

treatment [5, 51]. There are also many examples in cortex (see [14, 32, 38, 45, 46, 92] and [10, 13] for review) and cerebellum (see [47, 73] and [16] for review) where synaptic growth occurs with training, and such reorganization may even be required for learning [12].

Nonetheless, activity-dependent synapse stabilization has received comparatively little theoretical attention [50, 51, 60]. A good discussion of the strengths and weaknesses of the use of cortical rewiring can be found in [8].

The clusteron model of Mel addresses learning that can be accomplished solely by spatial rearrangement of synapses under the assumption of a nonlinear spatial summation of synaptic input [60]. Mel demonstrated that a single clusteron neuron could learn to respond more strongly to a set of training patterns than to randomly selected patterns of equal size (i.e., number of active synapses). Poirazi and Mel showed that synaptic rearrangement, in addition to allowing for increases in synaptic weights by way of redundant connections, can increase the capacity of the model [63]. In our simulations, with identical synaptic weights and 15% active synapses, synaptic rearrangement can enable a 1000 synapse clusteron to learn 20-25 training patterns for a range of learning thresholds and windows of integration. Our focus, however, was to study the dynamics of learning in the clusteron and some derivative models. We demonstrated how the choice of the learning threshold determined the speed and strength of learning, both through numerical simulations and analytically in a reduced model of the clusteron. The distributions of synaptic activations are affected by the density of the patterns, as well as the radius of integration. However, when

considering the relative location of the distribution to the learning threshold as we suggest, the trends we show hold for patterns of high and low density, as well as with large and small regions of integration. Finally, by modifying and extending the basic protocol, we showed how a clusteron-like rule can be used to learn sequences of patterns or different features of the training set.

In the clusteron, spatially clustered synapses must result in a nonlinear increase in overall efficacy. Poirazi and Mel argued that the storage capacity of a system making use of this nonlinearity increases dramatically [63]. Nonlinear information processing in the dendritic tree is well substantiated and has been shown to be responsible for several behaviorally relevant computations. Euler *et al.* have shown the earliest known location for direction selectivity in the mammalian retina occurs as a result of dendritic morphology of starburst amacrine cells [25]. Likewise, the precise structure of the dendritic arbor of the motion-sensitive neurons of the lobula plate of the blowfly correlates with their preferred direction of motion [48]. Furthermore, the dendritic tree of those neurons acts to filter the many phase-shifted inputs, representing the same signal, to generate the common output that is sensitive to the overall motion of the visual field [74]. Detection of an object on a collision course with an insect has also been attributed to nonlinear dendritic computation [28]. Even in networks where the input and output are more abstract, such as the hippocampus, or neocortex, nonlinear dendritic summation appears to be a prominent feature [49, 54].

While the clusteron was meant to model a single neuron, it provides a basis for describing a layer of neurons over which reshuffling of synaptic contacts could occur.

It is well documented that synaptic growth occurs, even in the mature human brain, as it learns a motor skill [45,86]. Inducible changes in the number of spines [18,21,83], “maturation” of the spines [37,58], and rapid filopodial growth in acute slices in real-time have been observed [57,84]. Furthermore, Stepanyants *et al.* estimated a so-called “filling-fraction” for various regions of the brain, to estimate the number of possible synaptic contacts that could be made by a short filopodial outgrowth. Their conclusion is that a large contribution to network remodeling could be made solely by growth of new spines [77]. Thus, while the clusteron seems to make use of fairly drastic reshuffling, in the context of a full layer of postsynaptic sites, small filopodial movements may be expected to find adequate sites.

Our simulations predict further properties of systems that learn by structural modifications. We can expect that the speed of memory acquisition and the strength and stability of the memory, will be strictly dependent on the probability of making an appropriate synaptic connection. Furthermore, the mechanism by which those dependencies emerge is clearly shown by the distribution of synaptic activations. Beyond making predictions about when a neuron can and cannot learn, we show that feature detection and sequence learning are both explained by the relationship of the synaptic activations to the value of the learning threshold. We therefore confirm the utility of our finding to explain the behavior of the model in these two tasks.

We believe that the present method of analysis and these findings are generalizable beyond the study of the clusteron and its derivative models. It seems that any model that allows for the formation of new synapses, either by axonal growth, or

by activation of silent synapses, could be described in a similar way. Any time a new synapse forms, it can either be suitable or unsuitable for the purposes it has to fulfill. The probability of either event will determine how well and how fast the system can learn. In our case, the suitability of a synapse was determined by its cluster membership, and approximately followed a normal distribution. This is only one possible measure of suitability. An alternative measure of suitability of a new synapse would be simply whether or not the synapse has formed on the correct cell. The distribution of synaptic suitability would then be the proportion of synapses that grow to the correct cell, or to the wrong cell, where synapse stabilization is allowed only for the correct cell. The rate of learning would then be proportional to the probability of synapse formation on the correct cell.

A simple analogy for this generalized model can be made to classical conditioning paradigms, where the location of the learning threshold is a measure of the difficulty of the task, which in this case would be related to the saliency of the conditioned stimulus (CS). Classic work of Pavlov and others have shown the distinctiveness of the CS to be critical in determining the rate of acquisition of the CS-US relationship (summarized in [75]).

Perhaps the most accessible example of learning that follows this relationship may be in the acquisition of a complex sensory or motor skill, such as learning a new language, or learning to play an instrument. This form of learning is fairly slow and allows time for the structural changes that our model utilizes [45,86]. The magnitude

of the learned response at the cellular level is difficult to measure, but we would argue that the stability of a learned response over time would be proportional to what we define as the magnitude of the response, since any degree of unguided structural remodeling would take longer to disrupt a larger response. It is well known that the length of time spent practicing a skill leads to a longer duration of memory retention [2,4]. Interestingly, analogous to the gradual increase of our learning threshold, skill learning also benefits from making incremental increases to the difficulty of the task [68,87]. Young pianists learn simple songs before Rachmaninoff, as our models benefit from a learning threshold which is initially low and raised gradually during training.

Chapter 3

Patterns of back-propagation in live neurons

While learning by structural rearrangement discussed in the last chapter has many interesting features, it is not the only means available to the cell to change its responsiveness to specific inputs. Cells communicate with each other through synaptic contacts [44,76]. The changes in strength of these synaptic connections are thought to be another mechanism by which learning and memory occurs at the cellular level [7,15]. Understanding the rules that govern these changes is therefore fundamental to understanding how the nervous system functions.

The pruning and generation of synapses may mediate coarse changes in neuronal interactions. On the other hand, changes in synaptic strength occur on timescales that are much shorter than those of growth and pruning. Conceivably, learning mediated by synaptic weight changes could even mimic the effect of synaptic generation by increasing the impact of synaptic connections that have negligible impact at the start. Similarly, a significant reduction in synaptic strength could in effect be similar to pruning. Given a sufficiently high connectivity between cells, all the benefits of growth and pruning could be attained through synaptic weight changes. Moreover,

this can be achieved far more rapidly than the time necessary to generate or prune synaptic contacts. The combination of methods may reflect the balance between the ability to learn quickly and the metabolic demands of having dendritic arbors large enough to have such high connectivity.

Our experiments examine the activity of the cell under stimulation patterns that induce changes in synaptic strength. To examine the effects of such stimulation, we need to measure the activity at the sub-cellular level across a large portion of the cell. This is not a simple task, considering that the cell spans hundreds of micrometers in all directions, while the size of the dendrites, and other structures of interest, can be less than one μm in diameter. Furthermore, the signals frequently occur on timescales less than 1ms. Therefore, the task requires sub millisecond sampling rates to achieve good temporal resolution, and sub micron spatial resolution. The small size of the structures and rapid timescale of neuronal activity are the main obstacles to a detailed characterization of neural activity.

Fluorescence microscopy provides a way of examining the structure and function of living cells that are inaccessible to direct electrical recording. Optical recordings are typically done by including a contrast agent that translates a biological or biochemical signal of interest into a measurable one (for reviews see [70, 71]). Such is the case for calcium measurements. Calcium-sensitive dyes are typically fluorescent molecules that bind calcium ions, inducing a structural change in the dye molecule. This structural change then changes the excitation or emission of photons from the molecule, allowing the change in calcium to result in a change in the amount of

emitted light, which can then be quantified by a suitable detector.

There are several technical challenges in performing these kinds of experiments. First, optical recordings have traditionally been hindered by problems with light scattering. The structures of interest are frequently buried deep within the brain slice. This makes imaging using single photon fluorescence techniques difficult because a lot of excitation light is lost before reaching the preparation, and emitted light is redirected causing image quality to suffer. Furthermore, obtaining a good optical signal is often limited by the amount of light the experimenter can apply to the preparation without damaging the tissue. Finally, and most importantly, neurons are inherently three-dimensional structures, and can extend hundreds of micrometers in the axial direction. A standard microscopes setup is limited to viewing only one focal plane at a time. The need to physically move the objective lens or insert an additional lens in order to scan in the axial dimension severely limits the time resolution that can be obtained.

A number of labs, including ours, have been addressing these problems in several steps. First, the development of multiphoton fluorescence microscopy partially alleviated problems stemming from light scattering [17]. Multiphoton microscopy makes use of very long wavelengths (near infrared) which penetrate the scattering media better than the light used in single photon fluorescence. The fact that *two* photons are required to excite the fluorophore, within a very short time, means that fluorophore excitation is proportional to the *square* of the incident light intensity, $f = \alpha_{2p}I^2$. In single photon microscopy, the fluorescence is proportional to simply the light intensity

itself, $f = \alpha_{1p}I$. The nonlinear relationship of multiphoton fluorescence sharpens the focal spot, decreasing the volume of excited fluorophore. As a consequence, signals emerge from a very small region, approaching the sub-micron theoretical diffraction limit, at the focal spot. Standard confocal imaging requires a pinhole to select light only from the focal plane, which decreases the signal by rejecting some good photons. The small excitation volume used in multiphoton microscopy obviates the need for such a pinhole and allows for the collection of more of the emitted light. As a result of these improvements, experimenters have access to tools that are capable of imaging small neuronal dendrites, deep in neuronal tissue, with far reduced photodamage. This method addresses one of the key challenges in these experiments, allowing us to quantify fluorescence from the fine scale structures of the neurons, with high quality images and little photodamage.

Additional improvements had to be made, however, to obtain the temporal resolution required to record the fast signals of interest at multiple locations on the cell. Existing methods of laser scanning typically use mirrors mounted on galvanometers, which are used to scan the preparation. Our instrument uses a completely different method, acousto-optical deflectors (AODs), to deflect the laser. This allows for a much faster repositioning of the laser by large angles. Interestingly, with a change in the input to the AODs, which will be described in detail below, the laser collimation can also be changed. This allows for scanning above or below the natural focal plane of the objective lens without physical movement of the objective. Thus, three-dimensional scanning can be achieved at very high speed.

Using this three-dimensional scanner we examined the interplay between back-propagating action potentials (bAPs) and excitatory postsynaptic potentials (EPSPs). As mentioned in Chapter 1, the occurrence of bAP, preceded by an EPSP, is thought to amplify the bAP and result in a superlinear influx of calcium. This could be a mechanism of coincidence detection underlying Hebbian plasticity. We examined the role of dendritic sodium channels in the interaction between bAPs and EPSPs. We observe a large boosting of the calcium signal associated with a bAP by pairing with presynaptic EPSPs. This boosting is present whether the EPSP is paired with a single spike, or the tenth spike in a train. Somewhat unexpectedly, the boosting is restricted to the distal region of the dendrites. We propose an explanation for this, and discuss some of the implications for spike boosting as a mechanism for coincidence detection.

3.1 Methods

3.1.1 Electrophysiology & calcium measurement

Brain slices were acutely cut from the hippocampus of 4-7 week old Sprague-Dawley rats according to published procedures [67]. Briefly, our dissection solution consisted of (in mM): 110 choline chloride, 2.5 KCl, 1.25 NaH_2PO_4 , 25 NaHCO_3 , 10 Dextrose, 0.5 CaCl_2 , and 7.5 MgCl_2 . This cutting solution was partially frozen and used for intracardial perfusion, and for incubation while in the cutting chamber. Three hundred μm thick horizontal slices through the hippocampus were made on a Pelco

Vibratome 1000 Plus Sectioning System. After cutting, slices were transferred to our standard extracellular solution which was made up of (in mM): 125 NaCl, 2.5 KCl, 1.25 NaH₂PO₄, 25 NaHCO₃, 10 Dextrose, 2 CaCl₂, and 2 MgCl₂. For the incubation time, this extracellular solution was supplemented with 1.3mM Ascorbic Acid, and 3mM Na-Pyruvate. After a brief 15 minute incubation at 32 degrees C, the slices were left in the supplemented extracellular solution for at least one hour at room temperature. Cutting solution and extracellular solution was continuously aerated with a 95% O₂ / 5% CO₂.

CA1 pyramidal neurons were patched in the whole-cell configuration, and at least 20 minutes were allowed for diffusion of the fluorescent dyes. Internal solution consisted of (in mM): 120 K-gluconate, 20 KCl, 0.2 EGTA, 10 HEPES, 2 MgCl₂. This was supplemented daily with frozen aliquots of 4mM MgATP, 0.3mM Tris-GTP, and 7 mM phosphocreatine. Two hundred μ M Oregon Green Bapta-1, OGB-1, was included for calcium measurement, while 50 μ M Alexa-594 was included for structural visualization, also added from frozen aliquots.

In some experiments, a bipolar extracellular stimulating electrode was placed in stratum radiatum, approximately 20 - 50 μ m from the dendrite of the recorded cell. Stimulation protocols (described in Section 3.1.4) were applied, consisting of EPSPs, driven by the extracellular stimulator, and bAPs, initiated via the patch pipette at the soma.

3.1.2 Data acquisition and analysis

Data were acquired using custom software, RUF II, originally designed by R. Fink, and later modified and extended by G. D. Reddy, and myself. Time series data analysis and image processing was done in MATLAB using custom algorithms written by myself.

3.1.3 2D and 3D scanning

Structural imaging and functional recording are made using a custom 3D scanning microscope (See Figure 3.1(c) for schematic). Excitation light at 820nm transmitted to the preparation passing through a 750nm dichroic mirror (750DCXRU). This wavelength of light is capable of 2-photon excitation of both fluorophores present (OGB-1 and Alexa-594). Emitted light is reflected by the dichroic mirror and passes through either a 600:200 (HQ600/200) bandpass filter, for the structural signal, or a 550:100 (HQ550/100) bandpass filter to select the calcium-sensitive dye for the functional signal. All of these filters were purchased from Chroma Technology Corp, Bellows Falls, VT. As described in detail in the publication describing the development of the instrument, Reddy et al [67], our microscope is capable of rapid scanning in three dimensions. The setup includes four AODs, configured to deflect an incoming laser beam by an angle dictated by the frequency of the acoustic wave applied to the crystal.

Passing through an AOD, the angle of deflection, θ , of the laser light, is linearly related the frequency, f , and inversely proportional to the velocity, v , of the acoustic

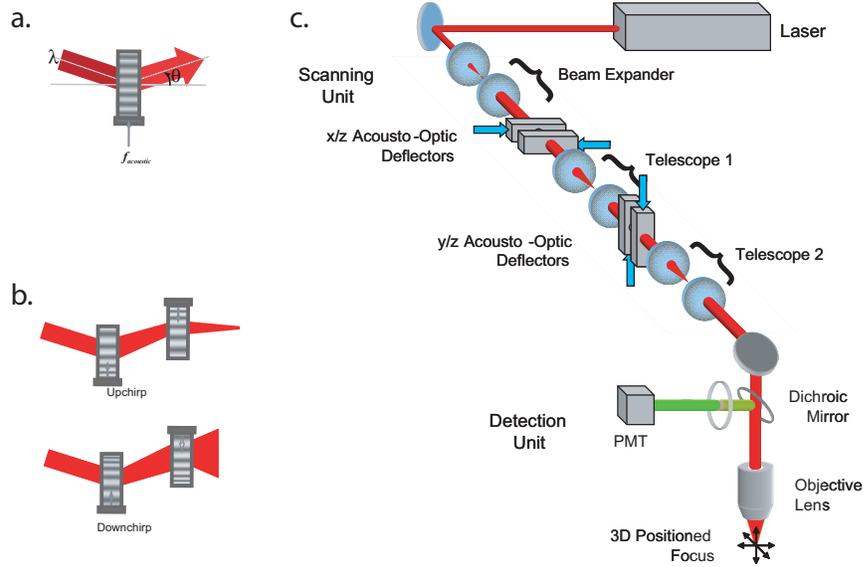


Figure 3.1: **Schematics of the optical setup.** (a) AODs deflect light by an angle θ , dependent on the frequency of the impinging acoustic wave (See Equation 3.1) (b) Providing a pair of counter-oriented AODs with linearly chirped acoustic waves deflects the laser by a range of angles. This results in a angular deflection of the beam, described by Equation 3.2, as well as a change in collimation of the beam, equivalent to a lens with focal length described by Equation 3.3. In practice, a 1:1 relay telescope is in place between the pair of AODs. (c) Schematic of the laser setup. Collimated laser light is first expanded and brought to the aperture of the first pair of AODs configured as explained in part (b). A 1:1 relay telescope transfers the image of the first pair of AODs onto the second pair. Light is guided to the preparation, first passing through a long-pass dichroic mirror, followed by the objective lens. Emitted light is directed by the dichroic through one of a set of changeable emission filters that selects the fluorophore of interest. Finally, light is collected by the PMT.

wave, as explained in the equation below:

$$\theta = \frac{\lambda f}{v}. \quad (3.1)$$

Note that the angle of deflection also depends on the frequency of the incident light, λ . For simple, two-dimensional scanning, similar to that achievable using standard galvanometer-based scanners only one pair of AODs is used. One AOD deflects the beam in the x -dimension (Figure 3.1(a)), while another perpendicular AOD deflects the beam in the y -dimension.

To scan in three dimensions, the second pair of AODs is also used, and all AODs are given linearly chirped inputs, rather than a constant acoustic frequency (See Figure 3.1(b)). Equation 3.2 describes the total angle of deflection, θ , for one pair of counter-oriented AODs. In this case, the AODs are given linearly chirped acoustic waves with center frequencies of f_{1c} and f_{2c} , which relates to the angle of refraction according to the equations below:

$$\theta = \frac{\lambda(f_{1c} - f_{2c})}{v} \quad (3.2)$$

$$F_{AOL} = \frac{v^2}{2\lambda\alpha}. \quad (3.3)$$

The term α corresponds to the slope of the acoustic chirps. The chirped frequencies results in a deflection of the beam by a range of angles, changing the degree of collimation of the beam. In effect, each pair of AODs, those oriented in the x , or those oriented in the y direction, create a cylindrical lens with a variable focal length, F_{AOL} , described above by Equation 3.3.

Because no physical movement of an objective lens or any mirrors is necessary, AOD-based scanners greatly improve the speed at which the laser can be redirected by a large angle. More importantly, our current configuration allows for a rapid repositioning of the focal spot three dimensions. The speed of movement is limited only by the aperture time, the length of time needed for a new frequency to fill the aperture of the AOD. As a consequence, there is no added time penalty for large deflections, or movements between focal planes. With standard galvanometer-based systems, the time to move to a new location increases with distance, and movements between focal planes require a separate technique.

A drawback to using AODs, on the other hand, is the large power losses, up to 30% from each AOD. The setup then requires a fairly high powered laser at the start of the optical path. Similarly, AODs impose significant temporal dispersion. This spreads out the laser pulses in time, resulting in a lower peak power. If power is a limiting factor, this can hinder experiments. Interestingly, the problem of spectral dispersion, where the wavelength dependence of diffraction from the AOD causes a smearing of the focal spot when scanning with two AODs does not appear in the four AOD configuration. The spectral dispersion by the first set of AODs, caused by the wavelength dependence of the angle of deflection (Equation 3.1), is compensated by the second set of AODs since they are oriented in opposite directions (Figure 3.1(b)).

3.1.4 Experimental protocol

As depicted in Figure 3.2(b), the stimulation protocol consisted of a 500ms baseline period, followed by ten action potentials at 10Hz, initiated at the soma. In some experiments, one of the 10 evoked action potentials is paired with a series of 3 synaptic EPSPs at 50Hz. These EPSPs were evoked starting at 50ms before the selected action potential by the bipolar extracellular stimulating electrode placed at 20–50 μ m from the dendrite of the recorded cell.

3.1.5 Calcium transient extraction

Functional traces were background subtracted and normalized for baseline fluorescence by calculating $\frac{\Delta F}{F}$. From these data, we extract the ten amplitudes of the calcium transients from the trace according to the following procedure. For each experiment, using the largest and cleanest traces of the cell, the decay phase after the tenth spike was fit to determine the best exponential time constant (See Figure 3.3(a)). Typically this was found to be about 600-1200ms. Thus, knowing the timing of each action potential, and the approximate shape of each transient, the peak amplitudes of those transients were estimated using a version of the Generalized Linear Model, GLM described next. Assuming that measurements were made at T points in time, our model has the form

$$D = \beta X + E, \quad D \in R^{1 \times T}, \beta \in R^{1 \times 10}, X \in R^{10 \times T}, E \in R^{1 \times T}. \quad (3.4)$$

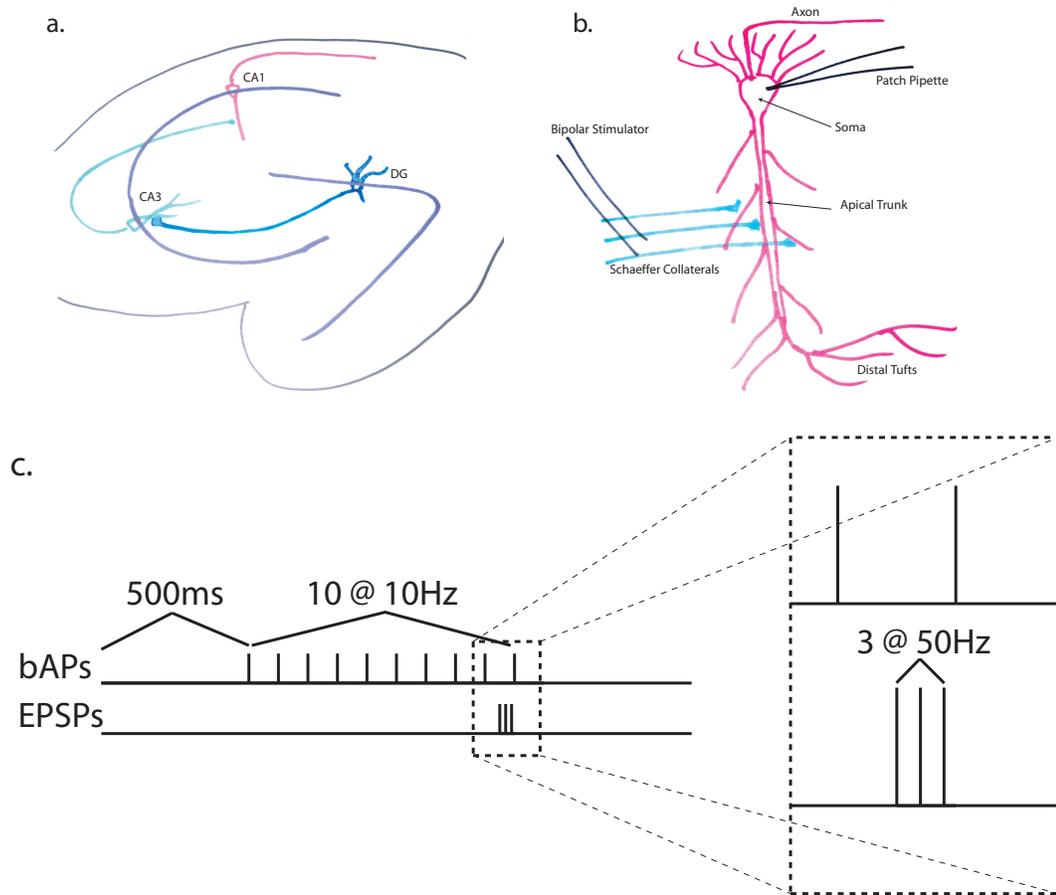


Figure 3.2: **Schematics of the experimental setup.** (a) A rough schematic of the hippocampus. Our experiment involves patch clamp recording of pyramidal cells in CA1, while stimulating the Schaffer Collateral inputs from CA3. (b) A closer view of the cell and position of electrodes. The patch clamp pipette is at the soma, while the synaptic stimulation is given in stratum radiatum, 20-50 μm from the dendrites. (c) The top plot represents the timing of stimulation through the somatic patch pipette to initiate bAPs. The bottom plot represents the stimulation via extracellular bipolar electrode. After a 500ms delay, 10 action potentials are initiated at the soma by a brief current injection. Preceding one of the spikes, either the 1st or the 10th, three synaptic stimulations at 50Hz are given, starting 50ms before the paired spike *i.e.* EPSPs triggered at $t - 50\text{ms}$, $t - 30\text{ms}$, and $t - 10\text{ms}$, with t the time of the paired spike.

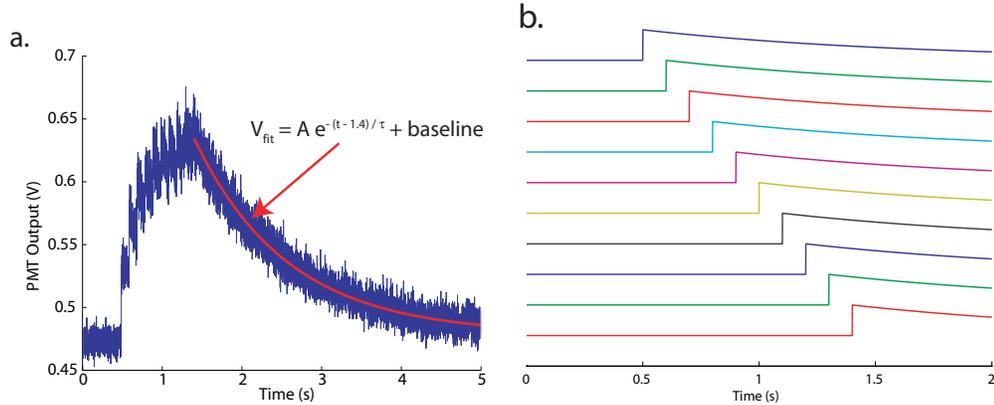


Figure 3.3: **Description of the GLM.** (a) To estimate the decay time constant of the fluorescence signal, the largest and cleanest signals of each cell were averaged. The decay to baseline after the last action potential was fit to a single exponential decay. The blue trace shows the average signal, as well as the best fit exponential in red. The parameter, τ , in the equation noted in the figure is used in the GLM analysis. The best fit timecourse in the example data shown above is 1.17s. (b) The model calcium transients corresponding to the matrix X . The τ corresponds to the best fit time constant from part (a), although there is a change in scale to better show the different transient onsets.

As a tenet of the GLM, we assume that the data vector, D , is a noise corrupted version of a sum of the the model calcium transients shifted in time to align them with the stimulation. Equation 3.4 represents this in matrix notation. The sum of functions is the product of a vector of coefficients, β , times the matrix X . The vector E represents the noise in the model resulting from the measurement or the nature of the observed process. The matrix X is analogous to the design matrix in some other fields, such as in the analysis of fMRI data [27, 41]. As shown in Figure 3.3(b), each row of the matrix is a model calcium transient induced by a single action potential.

The model shape is a function with an “instantaneous” risetime¹, and an exponential

¹The rise time of calcium after an action potential is very fast as is the binding of calcium to fluorophore. It is a simplification of this model that the risetime of the calcium signal if fast enough compared to our sampling frequency to be disregarded.

decay. More precisely, the i th row of the matrix X has the form

$$X_i = e^{\frac{-(t-s_i)}{\tau}} H(t - s_i),$$

where s_i is the time of the i th stimulation from the patch clamp pipette, and $H(t)$ denotes the Heaviside function.

To fit the model to data, we need to determine the vector of coefficients β . The value of those coefficients represent the amplitude of each model transient that makes up the fluorescent trace. Assuming that E is a vector of uncorrelated Gaussian random variables, the least squares estimate of the coefficients is given by

$$\beta = DX^T(XX^T)^{-1}. \tag{3.5}$$

Implicit in this this discussion is the assumption that the amplitude of the recorded signal is proportional to the amplitude of the underlying calcium transient. This assumption is discussed further in Section 3.2.

3.1.6 Spatial functional data analysis

The main challenge of the data analysis is to integrate measurements that are made from many points on the same filled cell. Both voltage and intracellular calcium are expected to change gradually over the length of the dendrite. Therefore, the concentration of bound fluorophore, and hence the true amount of light emitted from the recording sites, is expected to vary smoothly with path distance along the dendrite. This means the correlation between recording sites nearby can be exploited by functional data fitting in a further effort to increase the signal-to-noise ratio. I used

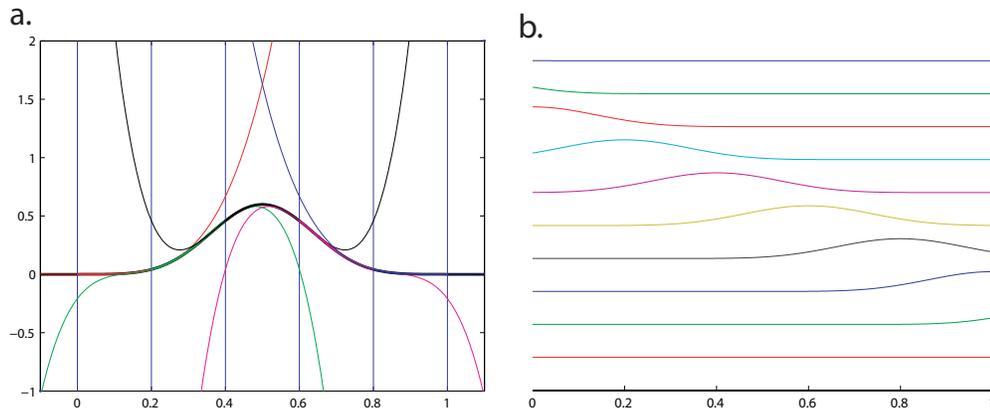


Figure 3.4: **Construction of a spline function.** (a) The function whose graph is represented by the heavy black line is fitted by splines. Fourth-order spline functions are defined piecewise between six knots, shown here as vertical blue lines. The red trace shows the polynomial which defines the function on the first interval between 0 and 0.2. Green, black, magenta, and blue define the function on the remaining four intervals. Outside the six knots, the function equals zero. (b) A simple example set of spline basis functions that could be used to fit data on the interval between 0 and 1. In our case, each function is a shifted version of all the others, with knots aligned. Fitting involves calculating the coefficients of these functions, which when summed, will approximate our data.

functional data analysis to better represent the smooth function assumed to underly the noise corrupted measurements (See Ref [66] for a good introduction to the theory of functional data analysis).

One-dimensional functional data analysis

The basis functions of our functional data analysis are fourth-order spline functions (See Figure 3.4(a)). Splines are piecewise polynomial functions, having six knots in our case, which are the points at which the polynomials are pieced together. Fourth-order splines are continuous up to the 3rd derivative, and smooth and continuous up to the 2nd derivative. A set of spline basis functions is defined on a domain matching

that of the data. Let B be a matrix an $M \times N$ matrix of spline basis functions as shown in Figure 3.4(b), where M is the number of basis functions, and N is the number of sites to evaluate them. Similarly, let C be the same size matrix representing the curvature of those basis functions, *i.e.* the second derivative of the spline functions. For ease of presentation, I define the $M \times M$ matrices, $J = B\omega B^T$, $K = CC^T$, and $L = J + \lambda K$. Here, ω is used to represent the confidence in each individual datapoint. It is an $N \times N$ diagonal matrix whose elements are the variance of the estimate for each datapoint. Also, λ is the so-called “stiffness” parameter, which essentially penalizes high curvature in determining the function that best fits the data.

The coefficients of the spline functions are calculated by

$$S = D\omega B^T L^{-1}, \quad (3.6)$$

where D is again a vector of datapoints. Note that if $\lambda = 0$, and $\omega = I_N$, the equation reduces to $S = DB^T(BB^T)^{-1}$. This is similar to the GLM solution for the coefficients of the model calcium transients (Compare to 3.5). After determining the coefficients, the smoothed function can be easily calculated by multiplying the coefficients by the matrix of basis functions,

$$\text{fit} = SB. \quad (3.7)$$

Prediction error for parameter validation

There are several free parameters when fitting a spline to the dataset. The first that needs to be set is the order of the spline, which is related to the order of polynomial

functions that comprise it. Cubic, third-order, splines are common and suitable for most purposes. Ours are fourth-order splines because we occasionally plot the second derivative of the spline, and the second derivative of a fourth-order spline is still a smooth function. The decision is not too critical however, since all spline basis functions beyond second-order have roughly the same shape and don't greatly affect the end result.

Two more parameters, however, are more critical since they can greatly affect the resulting fit. First is the knot interval, which is the distance between each of the knots. This controls how much space any given spline basis function will span. Having many sharply peaked functions, as a result of a small knot interval, allows the fit to match a rapidly changing function. However, the use of such function may lead to overfitting. On the other hand, splines defined with a larger knot interval will change much more slowly over the domain. Secondly, the stiffness parameter, λ , penalizes a curvature in the function, and pushes the calculated fit towards more slowly changing functions (See Figure 3.5(b)). It is imperative to have an objective way to set the parameters, beyond the attractiveness of the result.

Since we always have several replicates of each recording site under each experimental condition, we can validate our choice of those parameters by minimizing the *prediction error* of our fit. To calculate the prediction error of the functional fit, first, leave out one replicate of the data, and perform your smoothing based on the remaining replicates. The smoothed function is the “prediction” of the values of the missing replicate. The error is simply the sum of squared deviations between the

smoothed function and the missing replicate. The procedure is repeated by leaving each replicate out in turn. Finally the results are averaged. The value returned for this prediction error calculation is most useful when compared to similarly calculated values from fits with different parameters. In this way, the values for the knot interval, and spline stiffness, can be chosen according to which set of parameters best represents the data by minimizing the prediction error (See Figure 3.5(c)). If there were only one datapoint corrupted by Gaussian noise, the best estimate of the true underlying value is simply the sample mean of all the replicates. Functional data analysis takes advantage of the fact that nearby datapoints are correlated in their values. It thus improves on the estimate of the underlying value beyond what can be attained by using only individual sample means.

Branched functional data analysis

The previous section involved determining the best functional fit to a dataset defined on a simple one-dimensional domain, such as a subinterval of the real line. Our data, however, come from a branched domain. We present here an extension of traditional functional data analysis that makes use the specific correlational structure that smooth functions defined on a branched domain will impose.

As described above, B is an $M \times N$ matrix, where M is the number of basis functions, and N is the number of datapoints. This matrix contains the values of the spline functions at the locations of each datapoint. It is typically calculated as a function of location on the dendrite and the location of the knots. Another way to

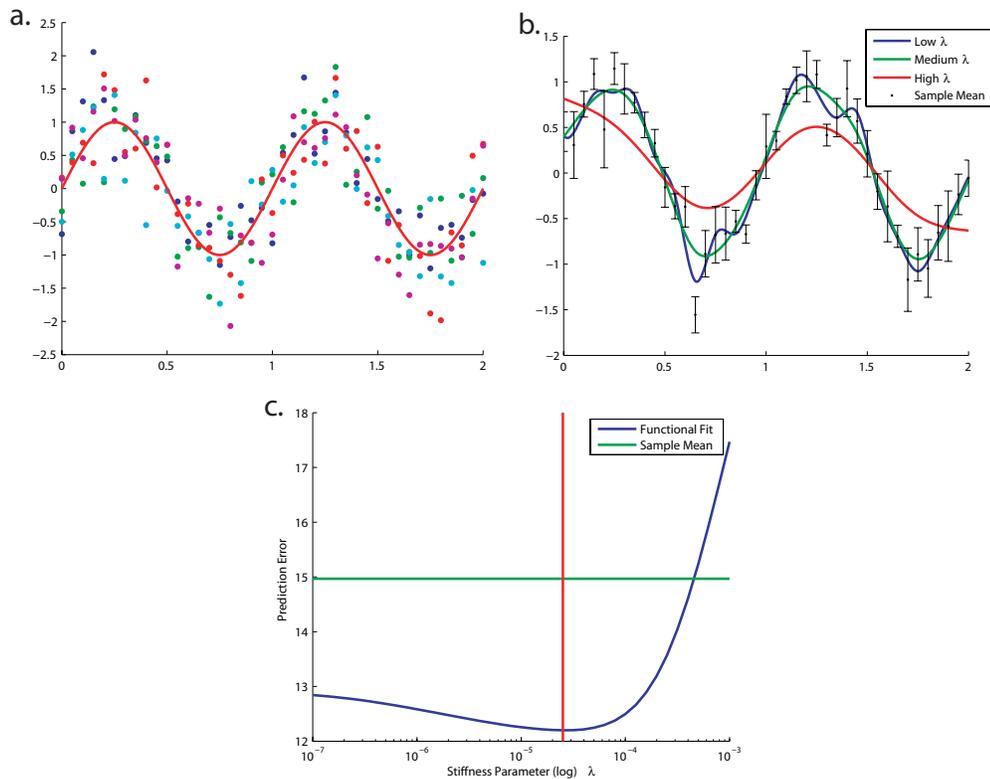


Figure 3.5: **Synthetic data and the basics of functional data analysis.** (a) Applying our analysis techniques to synthetic data allows us to test the effectiveness of the technique when the correct answer to the problem is known. The underlying function is a simple sine wave shown in red. Plotted also are five noisy “measurements” of the same function, as colored dots. (b) The effect of varying the stiffness parameter on the resulting functional fit. Blue, green, and red plots are functional fits with a low, medium, and high values for the λ . Black points represent the mean and standard error of the raw data. Note that the blue trace tends to follow each deviation in the underlying data, while the red trace is beginning to be too stiff to follow the data at all. When λ is very high, the functional fit becomes flat. (c) The choice of stiffness parameter is justified by minimizing the prediction error of the functional fit. Sweeping through a wide range of values for λ , the plot shows a clear minimum in prediction error around a value of $\lambda = 10^{-5}$, highlighted by the vertical red line. Compare this to the prediction error of the sample mean shown in green.

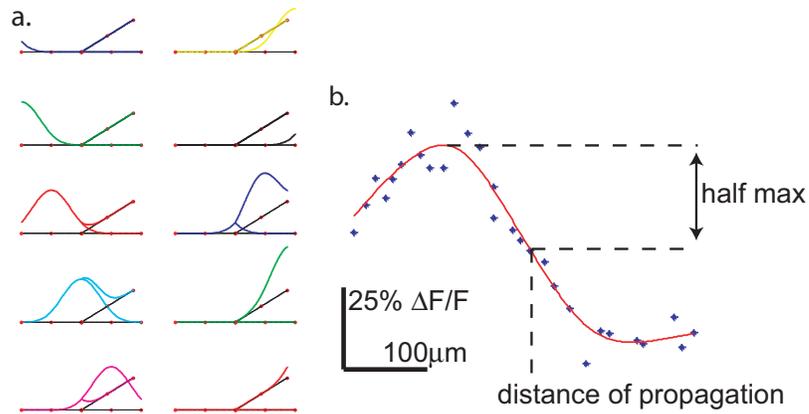


Figure 3.6: **Simple branched spline basis functions.** (a) Each basis function can be evaluated at any point on the tree by calculating the distance to the center of the spline. Notice how the function centered at the branch point (cyan) is equally represented in all branches, and functions near the branch point (red, blue, magenta) have some mass in all three segments of the tree. (b) Once there is a smooth functional representation of the data, a more reliable measure of the distance of spike propagation can be made. Here, the distance of propagation is defined as the distance at which the functional fit of the fluorescence values falls below half the maximum value attained.

think about it, which is easier to extend to branched structures, is to define a canonical spline basis function that is centered at zero. Evaluating the spline at the recording sites then involves calculating the distance from the recording site to the center of the spline function in question. A branched spline function centered near a branch point will have mass that extends down each branch since the distance from those points to the center of the spline is small. Figure 3.6(a) shows some simple branched spline functions that can be used to fit data on a domain with three branches. The branched functional fits of our real data are more complicated, as the domain of the function is determined by the branching pattern of the cell, but they always fit in the same mathematical framework. An additional normalization step is required for branched functions, since regions of the dendrite near the branch points have more spline basis

functions contributing than regions without branch points. All this entails is iterating through the columns of matrix B and normalizing the sum to 1.

Quantifying distance of propagation

Figure 3.6(b) illustrates of how we calculate the distance of propagation of a bAP. We define it as the point at which the functional fit drops to half of the maximal value attained on the tree. In summary, we have methods of extracting each individual calcium transient from the fluorescence traces, and functional data analysis to fit the entire cell's data at once. We therefore obtain a measure of the distance of propagation for all spikes in the train across the visible region of the cell.

3.2 Results

The ability to do fast scanning in three dimensions is a notable accomplishment by itself, and the subject of a standalone publication in *Nature Neuroscience* [67]. The scanner is capable of repositioning the laser to any location in the entire visible volume at high speed, limited only by the time at which a new acoustic frequency can fill the aperture of the AODs. This corresponds to the *settling time* of approximately $12\mu\text{s}$. We then keep the laser focus fixed at this location for $8\mu\text{s}$ for data collection, after which the laser is moved to a different spot. The entire cycle therefore takes only $20\mu\text{s}$, and a functional scan consisting of 50 arbitrarily positioned recording sites can be made at a 1kHz sampling rate.

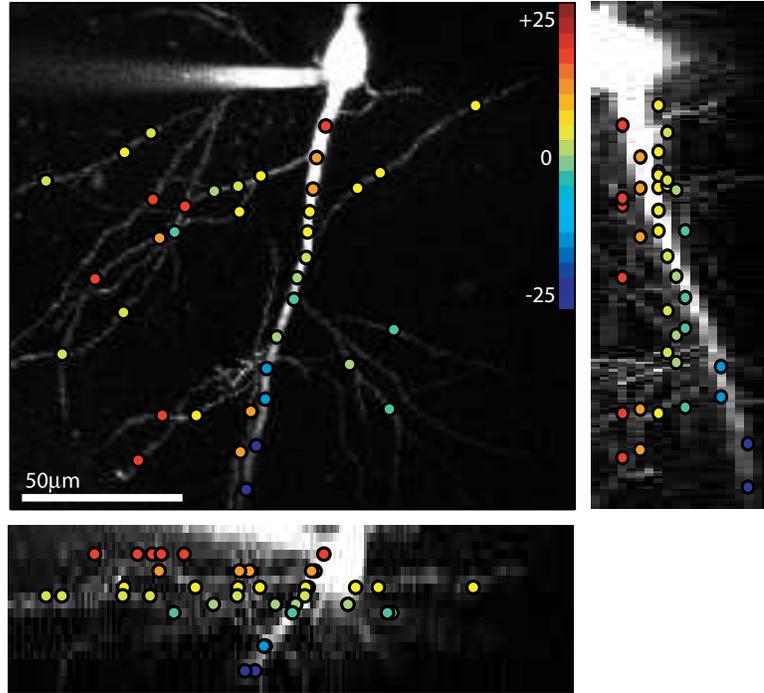


Figure 3.7: **3D scanning of live neurons.** The images above demonstrate the capabilities of the 3D scanner. These are maximal projection images from each of the three views, top-left is the z -projection, top-right is the x -projection, and the bottom is the y -projection. This example has 38 recording sites which are plotted on each image, color-coded for depth in the slice. It takes $8\mu\text{s}$ to scan a site and $12\mu\text{s}$ for the scanner to move between two sites. All points are covered in $760\mu\text{s}$ ($38\text{pts} * 20\mu\text{s}/\text{pt}$), before returning back to the first. The white scale bar is $50\mu\text{m}$.

3.2.1 Three-dimensional imaging of bAPs

Figure 3.7 shows three maximum projection images from a stack of data collected with this setup. All three main axis views are shown in an attempt to convey the three-dimensional nature of the dataset. After the image stack was made, 38 recording sites were selected, as shown on each of the images. The color of the marker represents the depth of field relative to the natural focal plane of the objective. As a proof of concept experiment, to measure the calcium transient associated with bAPs, I applied a simple stimulation protocol (3 bAPs @ 20Hz, 500ms delay). As shown in Figure 3.8,

the calcium transient associated with the stimulus is apparent (blue traces). After application of $150\mu\text{M}$ BaCl_2 , which has been shown to block the A-type potassium current [29], a large increase in the calcium response is clear in almost all traces from the apical trunk. This is consistent with the expected effect of the A-current on the shape of the action potential. As a very fast, depolarization-activated outward current, it is expected to decrease the amplitude of bAPs, and limit the range of back-propagation of a spike. This example illustrates the ability of the setup to measure the calcium signal induced by the bAPs, as well as its modulation by an A-type potassium channel blocker. It therefore serves as an important step in motivating and justifying the experiments that will be described next.

3.2.2 Linearity of the fluorescence response

In fluorescent imaging, there is always a concern about the linearity between the recorded optical measurement on the underlying signal. In the present case we are using the relative strengths of the optical signal to infer relations between the magnitudes of the recorded calcium concentrations. Therefore, in our case the concern about linearity is even greater, because we may be comparing the amplitudes of calcium transients from different baseline values. The fluorescence response is expected to follow an exponential binding curve. At high calcium concentrations, much less fluorophore is available. This saturation causes the measured calcium transients to decrease with dye occupancy. It is important that the response of our calcium dye remains approximately linear in order to compare the magnitude of different calcium

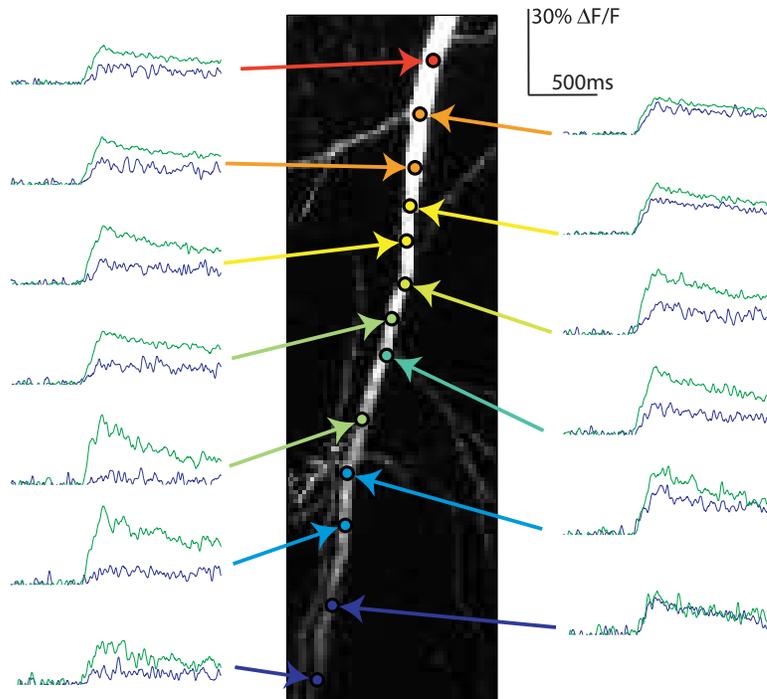


Figure 3.8: **K-channel dependence of bAPs.** For the cell shown in Figure 3.7, the effect of $150\mu\text{M}$ BaCl_2 bath application is shown for the 13 points on the apical trunk. The blue traces were recorded in control solution, while the green traces were recorded 20 minutes after bath application of BaCl_2 . The scale bar is shown in the top right of the figure. As expected, the blocking of the A current resulted in increased calcium response reflective of an increased depolarization due to bAPs.

transients.

Validation of our technique on this point is possible since for each cell, we have a number of recording sites at or near the soma. As verified by our simultaneous somatic recording, action potentials at the soma are reliable and have a consistent amplitude. This enables us to measure the fluorescence transient amplitudes at these sites near the soma to ensure that the amplitudes we extract via the GLM are of equal size. Figure 3.9(a) shows a sample recording site that is located near the soma. Note that the calcium transients extracted by the GLM are very consistent over the whole train in the site corresponding to the soma.

Some cells did exhibit a significant decrease in the amplitude of calcium transients with successive spikes, presumably due to dye saturation. These cells were not used in the comparison between different spikes in the train.

3.2.3 Pointwise analysis of fluorescence signals

Viewing the fluorescence traces from single recording sites as in Figure 3.9, a few confirmations can be made about the behavior of the bAPs. Firstly, as mentioned above, the fluorescent transient has a consistent amplitude when measured near the soma (Figure 3.9(a)), as expected from the amplitude of the action potentials there. Contrarily, the transients extracted from the dendrites do show reduction in amplitude later in the train (see Figures 3.9(b) and 3.9(c)). This is consistent with direct electrophysiological recordings which show that the amplitude of later spikes in a train recorded in the dendrite are smaller than the amplitude of the first. This is due

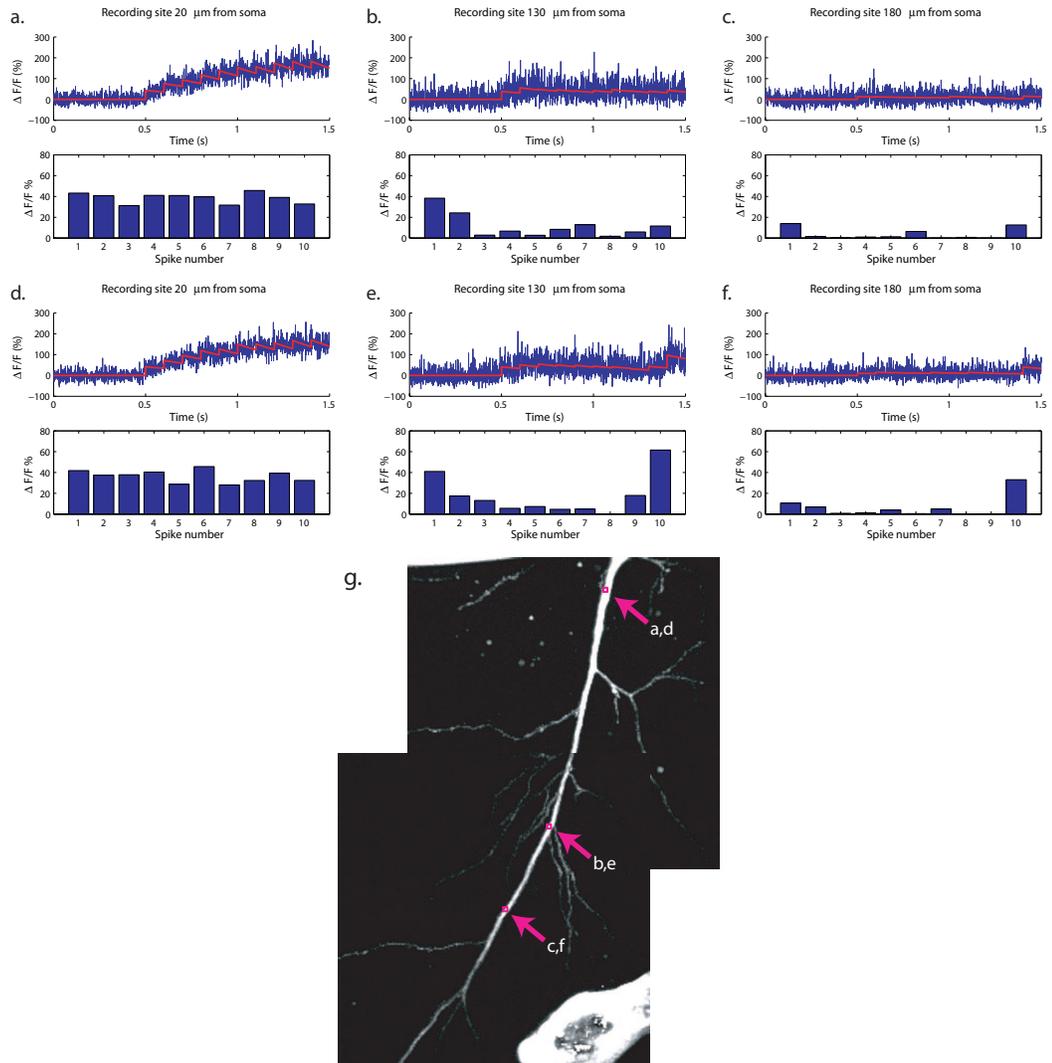


Figure 3.9: **Single point examples of fluorescence traces.** Left panels, (a,d), correspond to the proximal recording site, middle panels, (b,e) to the middle recording site, while the right panels (c,f) correspond to the distal recording site as shown in (g). The top pairs of plots are the response of the cell to ten action potentials initiated at the soma. In all cases, the blue trace is the raw data, and the superimposed red trace is the GLM fit. The bottom pairs correspond to ten action potentials, paired with three EPSPs before the tenth spike, as described in Figure 3.2(b). The soma is just above the image shown in (g). The bright object near the bottom of the image is the extracellular stimulating electrode.

to the slow recovery from inactivation of dendritic sodium channels [11]. Looking at the dendritic measurements in Figure 3.9(b) and 3.9(c), it is also apparent that the first transient appears to be decreasing in amplitude with distance, a phenomenon usually attributed to A-type potassium channels [36].

Additionally, we can observe significant boosting of the calcium signal after presynaptic stimulation just before the bAP. Compare 3.9(b) to 3.9(e) and 3.9(c) to 3.9(f) to see the changes in response with and without presynaptic stimulation paired with the tenth spike. Note that the amplitude of the tenth spike near the soma is similar between the two conditions.

3.2.4 Functional data analysis on the apical trunk

The data presented in the last section demonstrate the effectiveness of the GLM in extracting the amplitudes of the calcium transients and confirm several expectations from previous biological experiments. Our goal, however, is to generate a more complete representation of the activity level on the whole cell. Section 3.1.6 describes a method by which data recorded at different points of a simple one-dimensional structure can be fit to a smooth function. This method can be applied directly to our data. We first present the analysis of the data from the recording sites on the apical trunk of the dendritic tree. For the purposes of our analysis, this portion of the dendrite is represented by a line segment. The analysis of data on branched sections of the dendritic tree is shown subsequently.

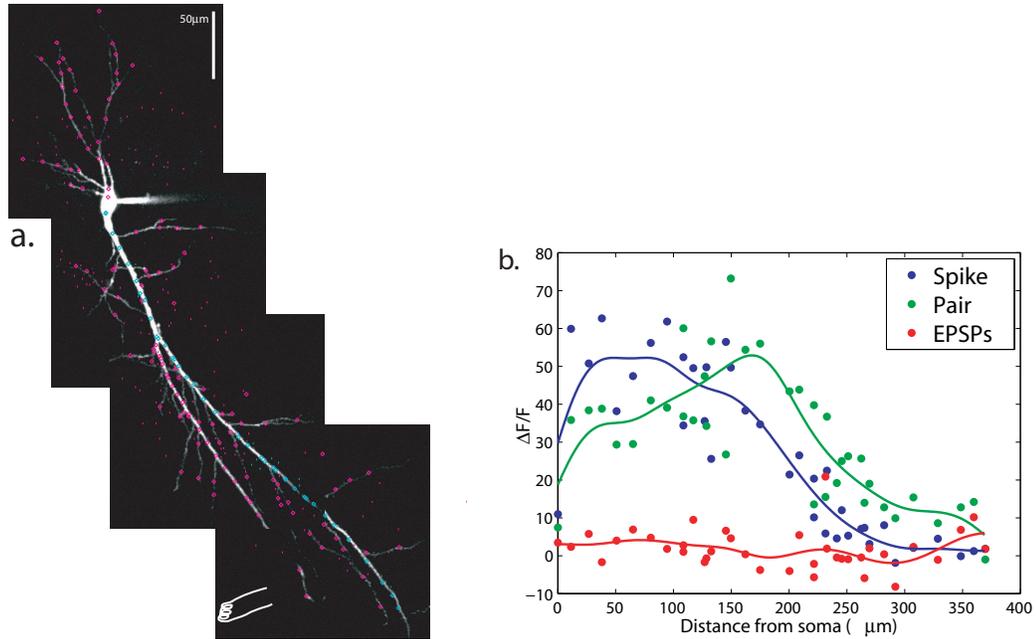


Figure 3.10: **Single spike propagation and boosting.** (a) Aligned maximum projection images from four recording regions of the experiment. The markers on the image are the locations of the recording sites. Cyan markers denote locations on the long dendrite used in the functional data analysis in panel (b). (b) Raw amplitudes, and functional fits for each of the three conditions of the experiment. Blue is the fluorescence signal resulting from an unpaired bAP, red is the signal from an unpaired presynaptic stimulation, and green is the signal from a bAP that is paired with presynaptic stimulation.

Figure 3.10 shows the recording locations and the amplitude of the calcium transient evoked by a single bAP. The markers on Figure 3.10(a) represent the recording locations during the stimulation. Figure 3.10(b) shows the raw amplitudes for the cyan recording locations from part (a), overlaid with the spline fit for each of the three conditions. The EPSPs that are evoked by extracellular stimulation have little effect on their own, but both the action potential and the pairing induce a large fluorescence change. Importantly, the signal from the action potential paired with extracellular stimulation extends significantly farther from the soma.

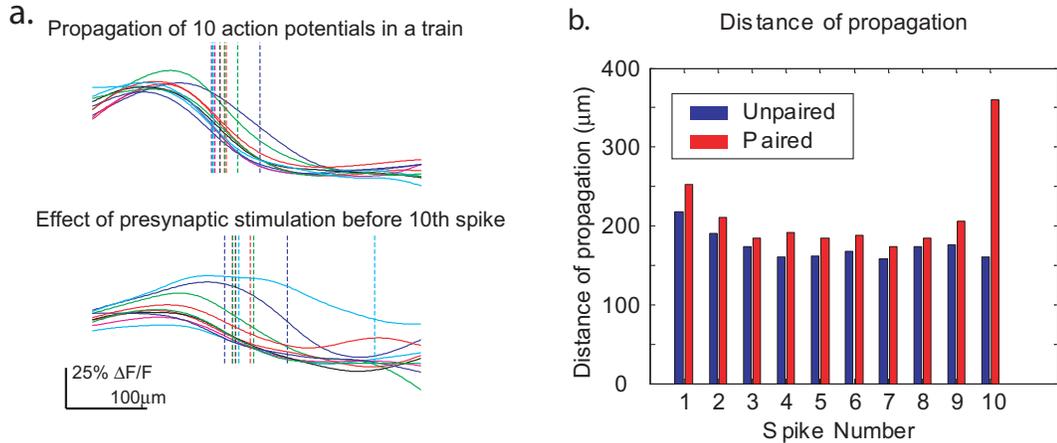


Figure 3.11: **Spike train propagation through the apical trunk.** (a) Top: A functional representation of each of the ten action potentials in a train, using all recording sites on the apical trunk. The dotted vertical lines represent estimates of the distance of propagation for each spike. Note that the blue and green lines correspond to the first and second action potentials in the train, which are known to propagate the furthest along the dendrites. (a) Bottom: The functional representation of ten action potentials, with presynaptic stimulation occurring before the tenth spike. Colors are the same as in the top plot. The cyan trace, corresponding to the tenth spike, extends further along the dendrite than even the first and second (blue and green as in the top plot). (b) Bar graph highlighting the distance of propagation of each of the spikes. Blue corresponds to trials without presynaptic stimulation, and red to trials with presynaptic stimulation just prior to the tenth bAP.

Figure 3.11 shows the one-dimensional fit to the data from recording sites on the apical trunk during a train of bAPs. The fit was obtained using functional data analysis, as explained in Section 3.1.6. Each calcium transient propagates backwards through the tree to varying degrees. In both cases in Figure 3.11(a), the response from the first spike (shown in blue) propagates farther than that of the second spike (green) which propagates farther than spikes 3–9. Only spike number 10, when paired with presynaptic stimulation propagates further (cyan trace, bottom plot). Figure 3.11(b) compares the distances of propagation between the two sets of trials.

It is interesting that in this case, as in the boosting experiment with a single bAP (See Figure 3.10(b)), the boosting is only apparent in the distal region of the cell. Some interesting implications of this observation will be discussed in Section 3.3.

3.2.5 Functional data analysis on the entire tree

The analysis presented above represent a great way to use functional data analysis to include more of the data in the measurement of the distance of propagation. On the other hand, a lot of data were left out, since we only analyzed data from the apical trunk. This is made obvious by the large number of recording sites in Figure 3.10(a) that were not included.

Using the branched spline fitting approach, described in Section 3.1.6, I fit the entire dataset at once, with smooth function defined on a domain that matched the branching pattern of the cell. The fit was obtained using splines adapted to the branching structure. Figure 3.12 shows the result of the fitting, where the amplitude of the first and tenth spikes are plotted with and without presynaptic pairing with the tenth spike. The bottom panels of Figure 3.12(a) show the difference in the amplitude of the calcium transients between the two conditions. As expected, the response of the first spike, which happens long before the presynaptic stimulation, is relatively unaffected. The difference plot in that case is relatively close to zero across the whole cell. On the other hand, the tenth spike, shown on the right plots is greatly affected on some branches, while unaffected on others. The result that the region of significant boosting is restricted to the distal branches is clear in this case as well.

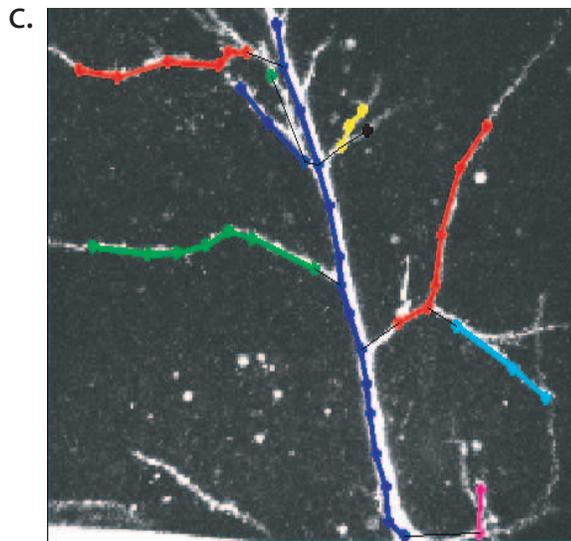
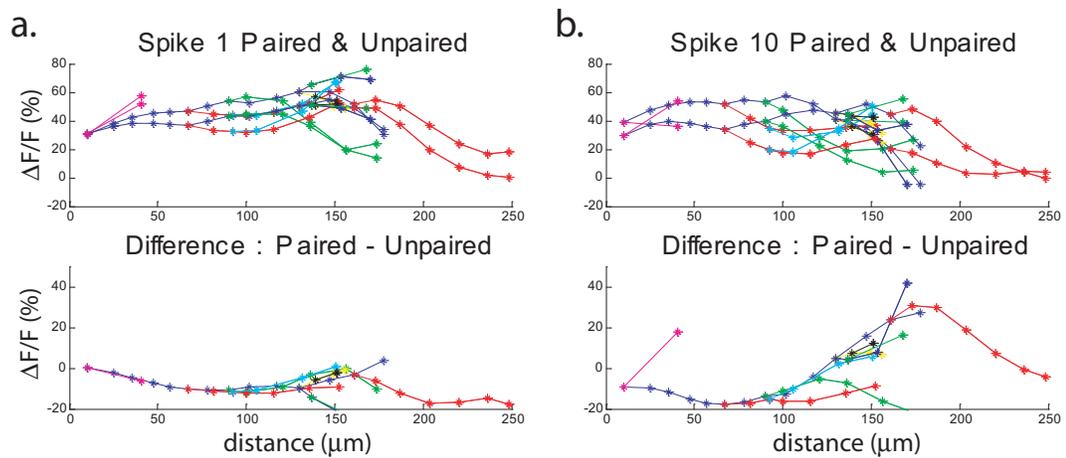


Figure 3.12: **Branched functional data analysis.** (a) The branched functional representation of the activity induced by the first action potential in the train, with and without EPSP pairing with the tenth spike. This is essentially a control condition, as the EPSP pairing with the tenth spike happens long after the first spike and should not effect it at all. Different dendritic branches are color coded according to the map image in (c). The bottom plot, the difference between the two functions, accentuates the effect of presynaptic pairing, which in this case is small. (b) The branched functional representation of the calcium response of the tenth spike, with and without pairing. Notice that the difference plot shows a large region of boosting of the calcium transient by the presynaptic pairing. Interestingly, this case shows what appears to be selective propagation down some branches (note the “failure” to propagate down the green branch, for instance). (c) An image showing the recording sites and color coded branches. The soma of this cell is just below the image, where the edge of the recording pipette can also be seen on the left.

3.3 Discussion

We have demonstrated the ability to measure the calcium transient associated with bAPs using 3D scanning microscopy. In particular, we used the technique to effectively estimate the extent of propagation of bAPs through a significant portion of the dendritic tree. The calcium wave displays many of the features expected from what is known about spike back-propagation. We show that the signal decreases in amplitude with distance from the soma, and decreases in amplitude with successive spikes in a train. Furthermore, the decreasing amplitudes of our calcium measurements are restricted to dendritic recording sites, as is the decrease in spike amplitude made from direct electrical recordings [11].

The new capabilities of our microscope instilled the need for new strategies for data analysis, that can take advantage of data generated from many locations at once. I presented the methodology for functional data analysis on simple one-dimensional domains, as well as on a branched structure. This technique, along with the GLM for time series analysis, goes a long way towards optimizing the estimate of the parameters of interest. An even more powerful model, which takes into account the nonlinearities imposed by fluorescent recording, as well as the noise profile of our instrument is presented in Chapter 4.

In the experiment presented above, we also observed the interaction of bAPs and EPSPs on the dendritic tree and have mapped the region in which the combination of stimuli produce a superlinear effect. An interesting point about the boosting of

the calcium signal that we observe (See Figures 3.10(b), 3.11(a), and 3.12(a)(b)) is that it is restricted to the distal regions of the dendrites (usually $> 150\mu\text{m}$).

One key component to how we understand the interaction of the two stimuli is in the distinction between spike amplification and resurrection of a failed spike. Both effects can change the amplitude of the action potential, and subsequently induce a measurable change in calcium fluorescence. Spike amplification by an EPSP would simply mean that presynaptic stimulation has inactivated the transient potassium channels, shifting the balance between excitatory and inhibitory currents. This changes the shape and amplitude of the spike by a degree proportional to the amount of potassium channels that were inactivated.

Resurrection of a failed spike, on the other hand, has some different properties. Acker and White [1] showed the effect of the changing concentrations of ionic conductances along the dendrite. They find that at some distance from the soma, the increasing concentration, and faster kinetics, of the A-type potassium channel results in a loss of the traveling wave attractor (TWA) for that membrane. This means that the conductances found far out on the dendrites are incapable of sustaining a regenerative action potential. An action potential propagating along this dendrite will reach a point beyond which only passive decremental propagation is possible. In this light, the effect of a presynaptic stimulation, as it inactivates the A-current, isn't merely a change in the balance of excitation and inhibition that amplifies the action potential. Instead, it changes the balance to restore the TWA, and allows for active propagation into regions where before there was only failure. This can result in a

much larger boosting of the spike since it moves across the bifurcation point where the TWA emerges.

Given the ability of the changing potassium conductances to allow or disallow active propagation through their impact on the TWA, it is easy to show a similar effect of changing sodium conductances. Our own simulations, data not shown, reveal that similar effect of loss of the TWA when sodium channel densities decrease. It is our hypothesis that with successive spikes in a train, due to the loss of sodium channels, the location of the bifurcation point of the TWA will migrate towards the soma.

The reason for making the distinction between spike amplification and spike resurrection is that the two mechanisms of spike boosting are likely to behave very differently. For example, the loss and gain of the TWA, will be restricted in space to the region of the dendrite close to the bifurcation point for the TWA, since a spike that has not yet failed cannot be resurrected, reemerging from the deep like Jonah from the belly of the whale. Actively propagating spikes, on the other hand, can still be amplified, since it is a variable change in spike amplitude, driven by any change in the balance of excitation and inhibition. There is also a difference in the gain of the boosting response to the two mechanisms. Even a small change in channel concentrations on a region of dendrite near the bifurcation point of the TWA can make a large difference in spike amplitude. On the other hand, a small change in channel concentrations will make only a small change in the amplitude of an actively propagating spike.

Our data show the region of the largest pairing induced calcium influx is in the

distal dendrites. This spatial patterning suggests our signal is dominated by the spike resurrection mechanism presented above. Without undue dependence on the theory of TWAs presented above, our data clearly show that the calcium response in the distal dendrites, from the bAP, is greatly influenced by presynaptic stimulation. Pairing with presynaptic stimulation may ensure back-propagation of the spike to the distal dendrites, and provide another form of coincidence detection for the cell. Multiple coexisting coincidence detectors has been proposed for corticostriatal STDP as well [26].

Chapter 4

Data analysis by sequential Monte Carlo particle filter

To review, we study the activity of neurons under different stimulation paradigms. Furthermore, our tools allow us to measure the level of activation of many different cellular compartments with high temporal resolution. This is accomplished with the use of fluorescent calcium indicators and a newly developed 3D scanning multiphoton microscope [67].

Our most basic analysis task involves the measurement of the amplitudes of the calcium transients evoked by a train of back-propagating action potentials. The simplest estimates of this amplitude are confounded by the low signal-to-noise ratio, and depending on the method used, can be biased. For example, measuring the amplitude of the transient as the maximum value during the transient [23,85] is biased towards higher values. It also suffers from the full extent of the recording noise, which dominates the measured signal when the light intensity is low (See Figure 4.1). On the other hand, measuring the amplitude as the average over an interval of time [3,6,69] during the transient is biased towards zero, since the calcium transient is not flat, but decays exponentially. As illustrated in Figure 4.1, this bias towards zero becomes

more problematic when measuring the amplitude of spikes in a train, where the signal has not yet decayed back to baseline. In studies with very low noise, such as when using a very slow detector, the signal can even be estimated by the first point after the stimulus [94].

The main focus of this chapter is to describe how to estimate the underlying signal, by using the information that is known about the processes by which our signal is produced and measured. These methods are then validated and characterized on synthetic data. The data analysis procedure makes use of our understanding of the kinetics of the binding reaction of Ca^{2+} and fluorophore, as well as the noise profile imparted by our instruments. This is accomplished through the implementation of a sequential Monte Carlo (SMC) particle filter, to find the most likely combination of parameters that produced a given data trace. This strategy has been used before in analyzing single channel traces [9], and more recently in fluorescence traces for extracting the spike times of a neuron [88].

The GLM model presented in Section 3.1.5 is a more sophisticated method than the traditional methods presented shown in Figure 4.1. It does take into account the shape of the expected waveform, and makes use of the whole dataset in an estimate, instead of only a subset of the available data. The limitations of the GLM, however, stem from the hidden assumptions that are made, such as the Gaussianity of the noise. I will show later that the detection noise from the PMT is highly non-Gaussian, especially at low light levels. The analytical tractability of the GLM also comes at a cost. The linearity of the model stems from the assumption that the solution is a

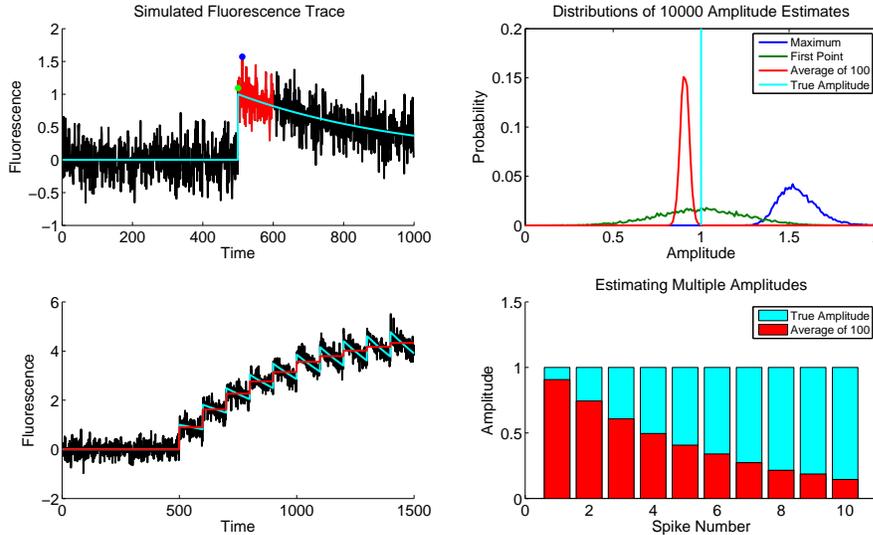


Figure 4.1: **Problems with traditional analysis techniques.** (a) Using synthetic data to compare the effectiveness of several different methods of estimating the amplitude of a simulated calcium transient (black trace). The calcium transient is modeled here as an instantaneous increase in fluorescence, followed by an exponential decay back towards baseline. The noise is assumed to be Gaussian with zero mean. Using only the first point (green dot), or the maximum value (blue dot), or an average over an interval of time (red section) are three ways to estimate the amplitude of a noise corrupted signal. The magenta trace is the noise-free signal. The cyan plots in all panels represents the true, noise-free signal. (b) shows the distribution of the 3 different estimates. Estimating the amplitude with a single point (green), with the maximum value (blue) and with an average of 100 points (red) show the effect of the different methods on the variance and bias of the estimate. (c) shows a simulated fluorescence response to a train of 10 action potentials, as well as the estimate of the fluorescence signal as an average of points during the transients (red). (d) shows those estimates as a function of position in the spike train. Note how estimating the amplitude of the transient by a simple average gets progressively worse later in the spike train.

linear combination of known functions.

The model-based procedure I present here makes no assumption of linearity, nor does it assume a Gaussian noise profile. Taking into account what is known about how calcium influx results in an increase in measured fluorescence, and how that fluorescence is measured by the detector, the most likely calcium influx is inferred from the measured fluorescence trace.

4.1 The model of the fluorescence signal

As mentioned above, the goal is to include knowledge of the system to best estimate the signal of interest, which is the degree of activation of each subcellular compartment. While voltage-sensitive fluorescent dyes measure the cell's membrane potential more directly, using the calcium signal as a proxy has a couple of practical advantages. First, calcium dyes offer a much larger change in fluorescence than voltage-sensitive dyes. Furthermore, a high affinity calcium dye, such as Oregon Green Bapta-1 (OGB-1), binds Ca^{2+} with rapid onset, but very slow offset (i.e. $\alpha \gg \beta$ in Figure 4.2). This allows for temporal integration of the signal, further increasing the signal-to-noise ratio of our estimate.

Figure 4.2 shows a simplified diagram of the chemical reaction that relates membrane potential to an increase in calcium concentration, and subsequently an increase in recorded fluorescence. Depolarization of the membrane acts to increase intracellular calcium concentration, $[\text{Ca}^{2+}]_i$, through the opening of voltage-gated calcium

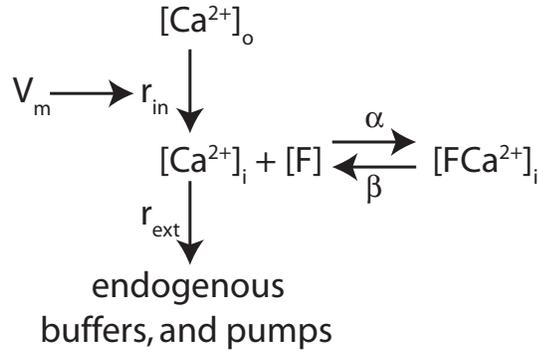


Figure 4.2: **The relationship between membrane potential and fluorescence.** Changes in membrane potential, V_m , affect the rate of Ca^{2+} influx into the cell via VGCCs. Intracellularly, $[Ca^{2+}]_i$ is sequestered by endogenous Ca^{2+} buffers, as well as by our optical probe, F . A change in the relative concentration of F and FCa^{2+} results in a measurable change in fluorescence.

channels, VGCCs. Under normal conditions, activity of VGCCs leads to a rapid increase in $[Ca^{2+}]_i$, which is then sequestered and extruded by a myriad of endogenous calcium buffers and membrane-bound calcium pumps. Our experimental manipulation makes this transient visible by including an additional calcium buffer in the cell. This buffer corresponds to the intracellular fluorophore, and is denoted by F . The final bound complex, FCa^{2+} , is 9 times more fluorescent than its unbound counterpart [33], providing an optical measure of electrical activity.

4.1.1 Fluorescence as a function of $[Ca^{2+}]_i$

The binding diagram given in Figure 4.2 provides the framework for knowing the way in which our signal of interest is transformed by our experimental setup. Changes in $[Ca^{2+}]_i$ are reflected in the resulting fluorescence changes, but the kinetics of the measured fluorescence don't necessarily match the kinetics of $[Ca^{2+}]_i$. In our case, with a

high affinity indicator, the onset of the fluorescence change is a good approximation to the onset of the Ca^{2+} signal, since the association rate of the two components, α , is very fast. The disassociation rate, β , however is much too slow to match the rate at which $[\text{Ca}^{2+}]_i$ levels return to baseline. Therefore, the resulting fluorescence is a “leaky integral” of the Ca^{2+} signal, rising quickly to match the rise of $[\text{Ca}^{2+}]_i$, but decaying back to baseline with a very slow time constant ($\approx 600\text{-}1200\text{ms}$). Equations 4.1 to 4.3 relate the three variables involved in the binding reaction explicitly,¹

$$\begin{aligned}
 [\text{Ca}^{2+}]_t &= [\text{Ca}^{2+}]_{t-1} + \sum_{j=0}^n A_j \delta_{t,t_j} + & (4.1) \\
 & dt \left(\beta [\text{FCa}^{2+}]_{t-1} - \alpha [\text{F}]_{t-1} [\text{Ca}^{2+}]_{t-1} \right) + \\
 & dt \left(r_{ext} \left([\text{Ca}^{2+}]_{base} - [\text{Ca}^{2+}]_{t-1} \right) \right) + \sqrt{dt} \chi_{Ca}
 \end{aligned}$$

$$[\text{FCa}^{2+}]_t = [\text{FCa}^{2+}]_{t-1} + dt \left(\alpha [\text{F}]_{t-1} [\text{Ca}^{2+}]_{t-1} - \beta [\text{FCa}^{2+}]_{t-1} \right) \quad (4.2)$$

$$[\text{F}]_t = [\text{F}]_{total} - [\text{FCa}^{2+}]_t. \quad (4.3)$$

The concentrations, $[\text{Ca}^{2+}]_t$, $[\text{FCa}^{2+}]_t$, and $[\text{F}]_t$ are calculated as functions of the concentrations and rates of change at the previous timestep, $t-1$. Here, r_{ext} represents a combined rate constant for all the endogenous mechanisms of Ca^{2+} sequestration, and $[\text{Ca}^{2+}]_{base}$ is the baseline Ca^{2+} level. The parameters, α and β are the rate constants for binding and unbinding to the fluorophore. Variation in the underlying calcium signal is represented by χ_{Ca} which follows a normal distribution with zero mean and standard deviation σ_{Ca} , that is $\mathcal{N}(0, \sigma_{Ca}^2)$. Therefore, calcium concentration in the model can rise and fall due to random fluctuations. The term r_{ext} will serve

¹Since our dye is intracellular, the measurements will only reflect changes in $[\text{Ca}^{2+}]_i$, and not $[\text{Ca}^{2+}]_o$, therefore the subscript will from now on refer to the time variable t .

to bring the calcium concentrations back towards baseline, which makes the underlying calcium concentration a mean-reverting Gaussian process. This process is a time discretization of an Ornstein-Uhlenbeck process. At the times of bAP initiation, the calcium concentration is allowed to increase by a variable amount, reflecting the unknown amount of calcium introduced by the bAP. In the model, a series of n parameters, A_j , represent the amplitudes of the ‘instantaneous’ rise in Ca^{2+} at the times of action potential initiation, which are denoted by t_j . Estimating the distribution of the most likely amplitudes is the main goal of this procedure.

Finally, light intensity I_t emitted from each recording location is related to the concentrations of free and bound fluorophore as

$$I_t = C_1(R_f[\text{FCa}^{2+}]_t + [\text{F}]_t) + C_2. \quad (4.4)$$

The scalar, R_f is the relative fluorescence of the bound fluorophore to free fluorophore ($R_f = 9$ [33]). The constant C_1 represents the proportionality constant for fluorophore concentration to light intensity, while C_2 is the signal offset resulting from autofluorescence.

4.1.2 Variability in PMT output

Now that the amplitude and timecourse of the signal of interest can be parameterized and predicted, additional knowledge of the detection system will also aid in uncovering the signal. Photomultiplier tubes, PMTs, are very sensitive light detectors, which amplify the light signal in several stages. Referring to Figure 4.3, an incoming photon

induces the release of a free electron at the photocathode. In the body of the PMT, a high voltage electrical field accelerates the electron towards a series of dynodes. At each dynode each incoming electron gives rise to a number of electrons which continue on to the next dynode. Therefore each dynode amplifies the signal multiplicatively. This results in an avalanche of electrons which grows at every stage.

The mean gain of the PMT is easily calculated. Given a PMT with k stages, where the gain per stage (the average number of electrons produced per incoming electron) is g , the mean total gain is equal to $G = g^k$. There is, however, uncertainty in the gain of the PMT, as the number of electrons produced at each dynode by an incoming electron is a random variable. The number of electrons emitted at each stage has been modeled previously by Tan [82], as a Poisson distributed random variable, with known mean, corresponding to the mean gain per stage. Under this assumption, a single electron at the first dynode, yields $p_g(1)$ electrons at stage two, where $p_g(x)$ represents the sum of x numbers drawn from a Poisson distribution with mean g . Subsequently, the number of electrons emerging from the second stage is $p_g(p_g(1))$, since *each* electron from stage 1 yields a group of electrons. Our eight stage PMT would be modeled by the eight level composite function, $p_g(p_g(p_g(p_g(p_g(p_g(p_g(p_g(1))))))))$.

Evaluating this composite function a number of times results in a distribution of gains that resembles that of a real PMT. However, due to the fact that the gain is not achieved by a simple additive process, the central limit theorem does not apply directly. As a result, the distribution of PMT gains at the output can be highly non-Gaussian, and depends on the incoming light intensity. At lower light levels, the

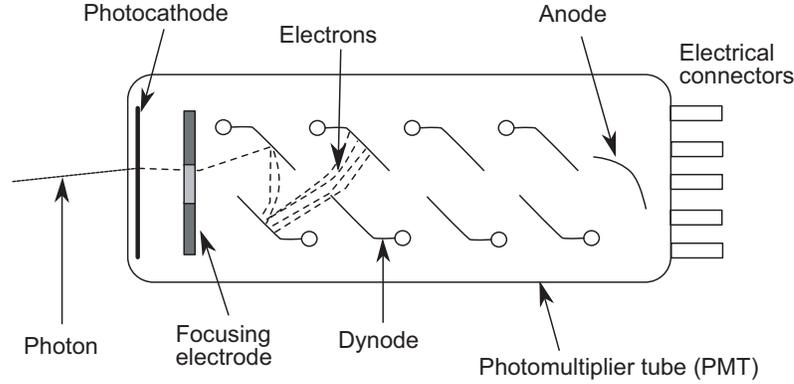


Figure 4.3: **Schematic of a PMT.** A photon enters the PMT and induces release of an electron at the photocathode. The electron continues towards a series of dynodes, which result in the amplification of the signal by a multiplicative factor at each stage. Public Domain image from [20].

distribution is highly skewed, approaching an exponential-like distribution at very low light levels. At higher light levels, the distribution becomes more normal, and becomes closely Gaussian at high light intensity. This change in the shape of the distribution of output voltages can be recorded directly in a PMT [22, 82], including our own as shown in Figure 4.4. Information about this gain distribution can be used to improve the estimate of the underlying signal.

4.1.3 PMT voltage as a function of fluorescence

To estimate the distribution of output voltages, i.e. the noise profile of the PMT, I collected many long data traces from a constant light source. This was done by imaging a simple fluorescent slide at different excitation intensities. Figure 4.4(a) shows four such data traces, corresponding to four different light intensities. The distribution of output voltages is approximated by a histogram of all points in the data traces, as shown in Figure 4.4(b). In all traces, the resulting distribution is well

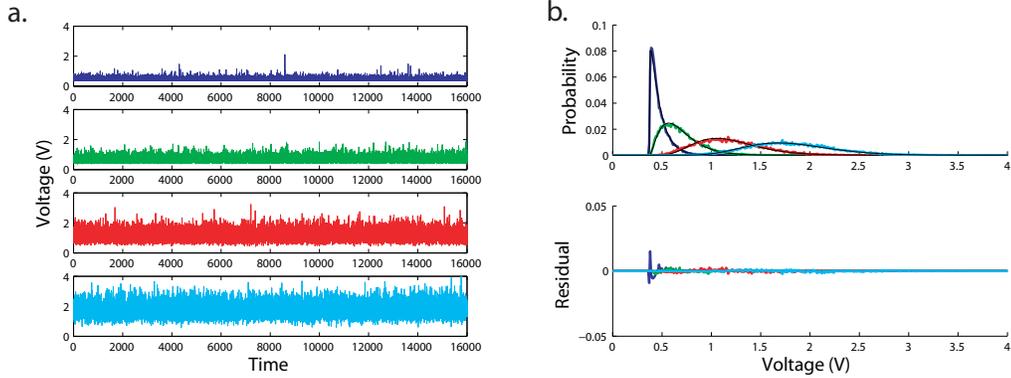


Figure 4.4: **Noise profile of PMT output.** (a) shows four voltage traces given a constant input from a fluorescent slide. The four plots correspond to increasing light intensities from top to bottom. (b) (top) shows the histograms generated from all points in each of the four traces in (a). Note that at low light intensity (blue), the distribution is highly skewed and appears exponential, while at high light intensity (cyan) the distribution becomes Gaussian. Superimposed on each trace in black is the gamma approximation given by the model. (b) (bottom) shows the residual not captured by the gamma approximation.

approximated by a gamma distribution. The two parameter family of gamma distributions reduces to the exponential distribution in certain limits, and the Gaussian distribution in others. This family has previously been used as an approximation to the gain distribution of PMTs [78].

Hundreds of noise traces like those shown in Figure 4.4(a) were used to build a model of the noise profile of our PMT. The gamma distributions that best fit each trace were found and define a function that relates the mean incoming voltage to the shape parameter, denoted here as A , of those gamma functions. Equation 4.5 was empirically fit (using Eureka version 0.79 [72]) for this purpose. The scale parameter B is then calculated as in Equation 4.6 to ensure that the resulting distribution has

the required mean value ($E[\gamma_{A,B}] = AB$).

$$A(I_t) = \frac{0.094}{0.112 + I_t^3} + 7.663I_t + 0.162I_t^2 - 0.476 \quad (4.5)$$

$$B(I_t) = \frac{I_t}{A(I_t)} \quad (4.6)$$

$$V_t = \gamma_{A,B} + C_3 \quad (4.7)$$

Here, $\gamma_{A,B}$ is a randomly distributed value which follows a gamma distribution parameterized by A and B , calculated from Equations 4.5 and 4.6. Note that values of V_t follow a shifted gamma distribution, on account of the voltage offset from the PMT, represented here as C_3 .

As mentioned above, the distribution of voltages from the PMT drastically changes shape with light intensity. A word of caution: Although the shape of the distribution is changing, and becoming highly skewed at low light levels, the amount of light emitted from the recording site is still best represented by the *mean* of the distribution, not the mode (peak value). That is to say, the amount of light emitted, our true parameter of interest, is proportional to the mean voltage at the PMT.

Consider a single fluorescent molecule emitting photons under constant excitation light. The emission of photons from that molecule can be thought of as a Poisson process, with events occurring at rate r . As such, the number of photons emitted in a given time interval, Δt , follows a Poisson distribution, with a mean $r\Delta t$. Now, assume that a second fluorescent molecule has arrived in the focal volume. The new rate of photon emission is $2r$, and the mean number of photons emitted in each time interval is $2r\Delta t$. Extrapolating further, the number of photons for n molecules of fluorescent dye

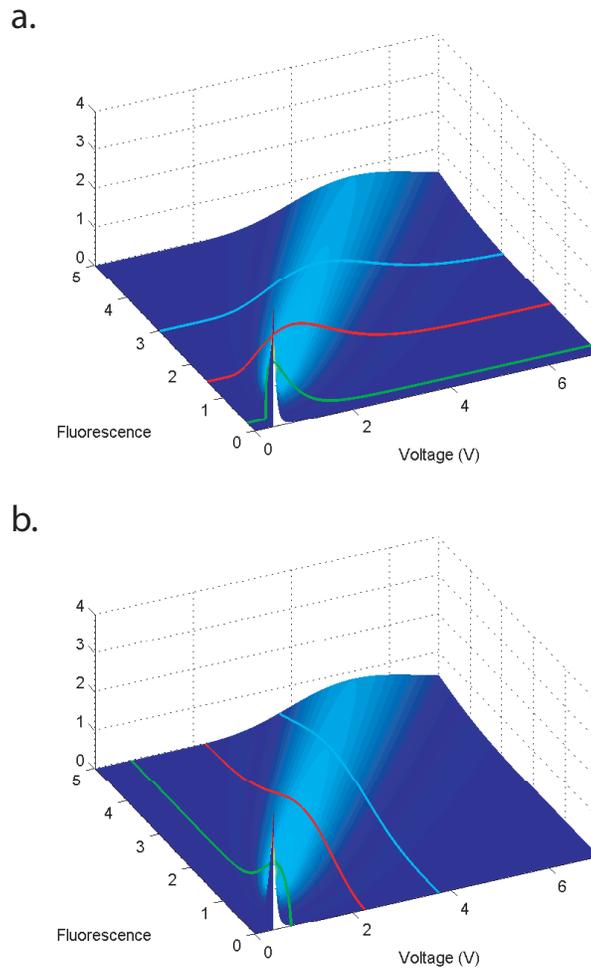


Figure 4.5: **Construction of, and inference from, the noise model.** (a) depicts the method by which a model of PMT noise was constructed from measurements of constant light sources. Highlighted are three distributions corresponding to low (green), medium (red), and high (cyan) light levels. Once the smooth function described in Equation 4.5 had been defined, the distribution of voltages from any light level, not just those used in the fitting, can be estimated. (b) illustrates the way in which estimates of light intensity are made from the model, assuming all light intensities are equally probable. Highlighted are three distributions corresponding to three different voltages measured from the PMT. These distributions are interpreted as the likelihood distribution of light intensities that produced a given voltage datapoint measured from the PMT.

is $nr\Delta t$. The shape of the Poisson distribution can also change with light intensity, from a highly skewed distribution at low light levels, to an approximately Gaussian distribution at high light levels. The shape of the distribution further depends on Δt , but the proportionality between the number of fluorescent molecules and the mean number of photons in a given time interval remains. In a similar way, the number of electrons emitted at a dynode of the PMT, is best described by the mean number of electrons it will emit, not the most likely number over a given time interval.

A key step in constructing the noise model is in definition of Equation 4.5. This relates the mean incoming light intensity to the parameters of the distribution of voltages that will emerge from the PMT. Once an equation has been found to describe this relationship, the distribution can be estimated for any possible value of incoming light, not just the ones that were recorded and used to fit the function. At that point, the problem can then be inverted using Bayes' Theorem. In our case, Bayes' theorem relates probability of light intensity, I_t , given a recorded value for V_t , in the following way:

$$P(I|V) = \frac{P(V|I)P(I)}{P(V)},$$

where $P(I)$ is the prior probability distribution for I_t . Here, $P(V)$ is a normalizing constant, which can also be written as $\sum_I P(V|I)P(I)$. In this analysis, we always assume a uniform prior distribution over I_t .

Thus, the voltage output of the PMT at time t , V_t , is an independent gamma distributed random variable with parameters which are functions of I_t . Constant C_3 represents the voltage offset of the PMT. Figure 4.5(a) shows the relationship of the

distribution of V_t to a given I_t , and depicts how the model of PMT noise is constructed. Highlighted are the distributions of noise at three constant light intensities. As will be discussed in more detail in the following sections, this noise model is used to calculate what likelihood distribution of I_t values produces a given V_t , as shown schematically in Figure 4.5(b). Highlighted are the likelihood distributions of I_t for three voltages measured from the PMT.

The assumption that each measurement is an *independent* random variable is validated by measuring the cross-correlation between noise traces collected in the same sweep. Recall that the laser focus sequentially visits each recording point, and then repeats the sequence until the end of the sweep. As such, if PMT samplings had temporal dependencies it would be reflected in the cross-correlation of successive points scanned. Since no cross-correlation was higher than 10^{-3} , the assumption of independence appears valid.

This section has described a slightly simplified model of how an increase in $[\text{Ca}^{2+}]_i$ leads to an increase in voltage output of the PMT. The model includes the major component which changes the timecourse of the signal due to the binding kinetics of our fluorophore, and the noise profile of the detector. We next use this model to infer what set of parameters most likely produced the recorded data traces.

4.2 Particle filter estimation of the calcium signal

We assume our system behaves according to the model presented above, and consists of a set of parameters we wish to estimate. We therefore have a set of hidden states

$$\mathbf{H}_t = \{[\text{FCa}^{2+}]_t, [\text{F}]_t, [\text{Ca}^{2+}]_t\},$$

along with a set of parameters,

$$\theta = \{A_j, [\text{F}]_{total}, r_{ext}, \sigma_{Ca}, [\text{Ca}^{2+}]_{base}\}.$$

These dictate the transition between possible hidden states, according to Equations 4.1–4.3. To constrain the model, we use the PMT output as the observation of the system at time t , $\mathbf{O}_t = V_t$. This results in a set of observations $\mathbf{O} = \mathbf{O}_{1:N}$. From the model of the system, detailed above, we define the observation distribution, which describes the probability of observation states, given a particular hidden state.

$$P(\mathbf{O}_t|\mathbf{H}_t) = P(V_t|I_t) = P(V_t|[\text{FCa}^{2+}]_t, [\text{F}]_t) \quad (4.8)$$

The probability of each observation is obtained using the gamma distribution, with parameters dictated by I_t through Equations 4.5 and 4.6. Our task, however, is to estimate the relative likelihood of hidden states at time t , given the entire set of observations, $P(\mathbf{H}_t|\mathbf{O})$.

The particle filter begins with the creation of a large number of particles, which are independent realizations of the model described above. The particle filter method allows us to estimate the likelihood distribution of hidden states, as they are constrained by the observation states, representing the experimentally recorded values.

More precisely, the distribution is approximated by a set of weighted particles as in:

$$P(\mathbf{H}_t | \mathbf{O}_{1:t}) \approx \sum_{i=1}^N w_t^i \delta(\mathbf{H}_t - \mathbf{H}_t^i), \quad (4.9)$$

where $\delta(\cdot)$ denotes a point mass at the hidden state defined by particle i . The weights are determined by:

$$w_t^i = \frac{P(\mathbf{O}_t | \mathbf{H}_t^i) P(\mathbf{H}_t^i | \mathbf{H}_{t-1}^i) w_{t-1}^i}{q(\mathbf{H}_t^i)}. \quad (4.10)$$

We use a simple version of the algorithm and assume that the sampling distribution, $q(\mathbf{H}_t^i)$, is equal to the transition distribution, $P(\mathbf{H}_t^i | \mathbf{H}_{t-1}^i)$. Therefore,

$$w_t^i = P(\mathbf{O}_t | \mathbf{H}_t^i) w_{t-1}^i. \quad (4.11)$$

Here the weight, w_t^i , of particle i , at time t , is calculated as the product of the weight of the particle at the previous timestep, and $P(\mathbf{O}_t | \mathbf{H}_t^i)$, which is the likelihood of the observation from particle i . In other words, the weight of the particle defined by the hidden parameters, \mathbf{H}_t^i is updated by the likelihood that this particular particle generated the voltage value measured at time t .

4.2.1 The SMC particle filter algorithm

The particle filter approach is based on the simulation of a large number of particles, all of which are potential candidates for the system that created the data trace. Although it would be very computationally intensive, one could, in a very simple way, create an enormous number of particles, calculate the likelihood of each in producing the recorded data, and find the particle with the highest likelihood. For data with

only one or two parameters this is a feasible approach. The curse of dimensionality eventually makes this approach impossible, however, because of the exponential increase in the number of particles with added dimensions.

Particle filter algorithms avoid such difficulties by attempting to ignore the large regions of parameter space that are very unlikely to have generated the data. This is achieved by periodic resampling from the set of particles: Particles with very low weight (and hence low likelihood) are erased. To keep the population of particles constant, new particles are generated from those with high weight. Such algorithms can achieve a good approximation of the distribution of interest with relatively small numbers of particles.

A basic implementation of the analysis method to estimate the entire set of hidden states, \mathbf{H}_t , as well as the total fluorophore concentration, and the amplitudes of the ten action potentials in a trace, is shown in the code block below.

- * Create N particles
 - * Randomly initialize parameters $\{[F]_{total}, A_j\}$
 - * Initialize particles with equal weight : $w_{t=1}^i = 1/N$
- * For each time point where V_t has been recorded
 - * Integrate Equations 4.1–4.3 up to that time
 - * Calculate $I_{i,t}$ predicted by each particle from Equation 4.4
 - * Calculate $P(V_t|I_{i,t})$ using the recorded value of V_t
 - * Update the weight of each particle : $w_t^i = w_{t-1}^i P(V_t|I_{i,t})$
 - * if($N_{eff} < N/2$) then resample the particles
 - * Randomly sample, with replacement, from the particles, weighted by w_t
 - * Reset particles to equal weight : $w_t^i = 1/N$
- * Interpret data
 - * Generate kernel density estimates of parameters
 - * Peak of KDE is most likely value

Here again, i indexes the particle, while t indexes the timestep. After the likelihoods are updated for each timestep, the *number of effective particles*, is computed as $N_{eff} = \frac{1}{\sum_i w_t^i}$. Notice that when all the weights, $w_t^i = 1/N$, then $N_{eff} = N$. Also, if all the weight was in one particle, *i.e.* $w_t^x = 1$ and all $w_t^{i \neq x} = 0$, then $N_{eff} = 1$.

Since the resampling step eliminates particles with bad estimates for the parameters, by the end of the algorithm, most of the particles will have a decent estimate of the value for each parameter. We use a weighted Gaussian kernel density estimate to approximate the probability density function of the model parameters from the discrete values contained in the population of particles. The most likely value is the one at the peak of the kernel density estimate.

For example, when iterating through each time step, at the time of an action potential, each particle has an estimate as to what the amplitude of the calcium transient is. Each particle predicts what the fluorescence intensity should be. The weight of each particle is updated by how likely the recorded voltage was produced by that particle's predicted intensity, $P(V_t|I_{i,t})$. Particles with unreasonable estimates of the amplitude of the calcium transient will suffer, because the measured voltage will have been an unlikely value to occur based on their predicted fluorescence. Particles that guess close to the value represented by the data will retain a relatively high weight. When the particles are resampled, they are randomly selected *with replacement*, weighted by w_t^i . This means that those with high weight have a good chance to continue, and a good chance to get duplicated, should they get selected more than once. Particles with low weight are probabilistically discarded in this resampling

step. At the end of the analysis, a histogram of the particles' values will estimate the likelihood distribution of that parameter. The Gaussian kernel density estimate that we use acts in a similar way, but generates a smooth function at the end.

4.2.2 Summary of parameters

A summary of model parameters retrieved from published sources, or calculated directly is shown in Table 4.1. Those parameters estimated by the Monte Carlo particle

Symbol	Units	Parameter	Value	Source
α	$M^{-1}s^{-1}$	OGB-1 association rate	$0.79 * 10^9$	[19]
β	s^{-1}	OGB-1 disassociation rate	178	[19]
R_f	-	Fluorescence ratio $\frac{FCa^{2+}}{F}$	9	[33]
C_2	V	Autofluorescence	varies	Calculated
C_3	V	Voltage offset of PMT	0.3724	Calculated

Table 4.1: Fixed model parameters

filter are shown in Table 4.2. Note that not all of those parameters need to be estimated independently from all the data traces, instead one high quality estimate was made, and a single parameter was used for all traces. Those parameters are labeled with an estimation frequency of *once*. Other parameters are estimated at every point, *i.e.* $[Ca^{2+}]_t$, $[FCa^{2+}]_t$, while the A_j are estimated only once per spike. The total fluorophore concentration, $[F]_{total}$, is not expected to change during the time of a single trace, and so a single estimate is made for the whole trace.

Estimation Frequency	Symbol	Units	Parameter
every timepoint	$[\text{Ca}^{2+}]_t$	M	Free calcium
every timepoint	$[\text{FCa}^{2+}]_t$	M	Bound complex
one value per spike	A_j	M	Calcium transient amplitudes
one value per site	$[\text{F}]_{total}$	M	Total fluorophore concentration
once	r_{ext}	s^{-1}	Rate of calcium clearance
once	σ_{Ca}	M s^{-1}	Noise in calcium concentration
once	C_1	$\frac{\text{V}}{\text{M}}$	Proportionality constant ($\text{F} \rightarrow \text{V}$)
once	$[\text{Ca}^{2+}]_{base}$	M	Baseline calcium concentration

Table 4.2: Monte Carlo estimated parameters

4.3 Performance with synthetic data

To test the effectiveness of our method, we characterize its performance with a synthetic dataset, which is designed to reflect the statistical structure of our recordings. This approach is standard when comparing two parameter estimation techniques. We simply compare the estimates obtained by different approaches to the parameters used in generating the synthetic data.

First, the model is instantiated, the parameters of which are unknown to the SMC particle filter which will be used in the estimation. In this example, values for the parameters A_j and F_{total} are set, and the initial conditions for variables $[\text{F}]_t$, $[\text{FCa}^{2+}]_t$, are chosen. Equations 4.1–4.3 are recursively evaluated to generate values for each variable at each time t . At that point, the entire vector of light intensity over time can be calculated by Equation 4.4. Figure 4.6(a) shows the result of this process. Among the hidden set of values are: free calcium (top plot), bound fluorophore + calcium (second plot), free fluorophore (third plot), and light intensity (bottom plot). These values are part of the hidden set of variables, described above as \mathbf{H}_t . As such

the analysis does not have access to these values.

We highlight some important features of the traces in Figure 4.6(a), since they illustrate how the model works. First, the effect of calcium influx at the times of the bAPs can be clearly seen in all the plots. The amplitudes of those calcium transients will be estimated by the algorithm as well. It is important to note that the plot of free calcium (top plot) shows these transients as well, and their shape closely matches that of the other plots. This may be somewhat unexpected since it has been previously stated that the decay of a normal calcium transient is very fast, much faster than the decay to baseline of the fluorophore. It can be shown, however, that integration of Equations 4.1–4.3, without fluorophore, *i.e.* $[F]_{total} = 0$, does in fact yield a trace with very rapid transients of free calcium. When fluorophore is included, however, the amplitude of those calcium transients decreases greatly, and their decay time constant lengthens. Another point to make regarding Figure 4.6(a) is in regard to the plot of free fluorophore (third plot). As expected, the plot in this case is an inverted form of the plot of bound fluorophore (second plot), since the amount of free fluorophore decreases when calcium binds to it. Interestingly, as this value approaches zero, the reduction in the amount of free fluorophore necessarily leads to a decrease in the sensitivity of the fluorophore to changes in calcium. This occurs exactly when the calcium dye saturates, and is the major source of nonlinear effects in optical measurements.

Once the vector of fluorescence values, I_t , has been determined, an instantiation of a possible voltage trace, V_t , can be generated. Shown in Figure 4.6(b)(top), the

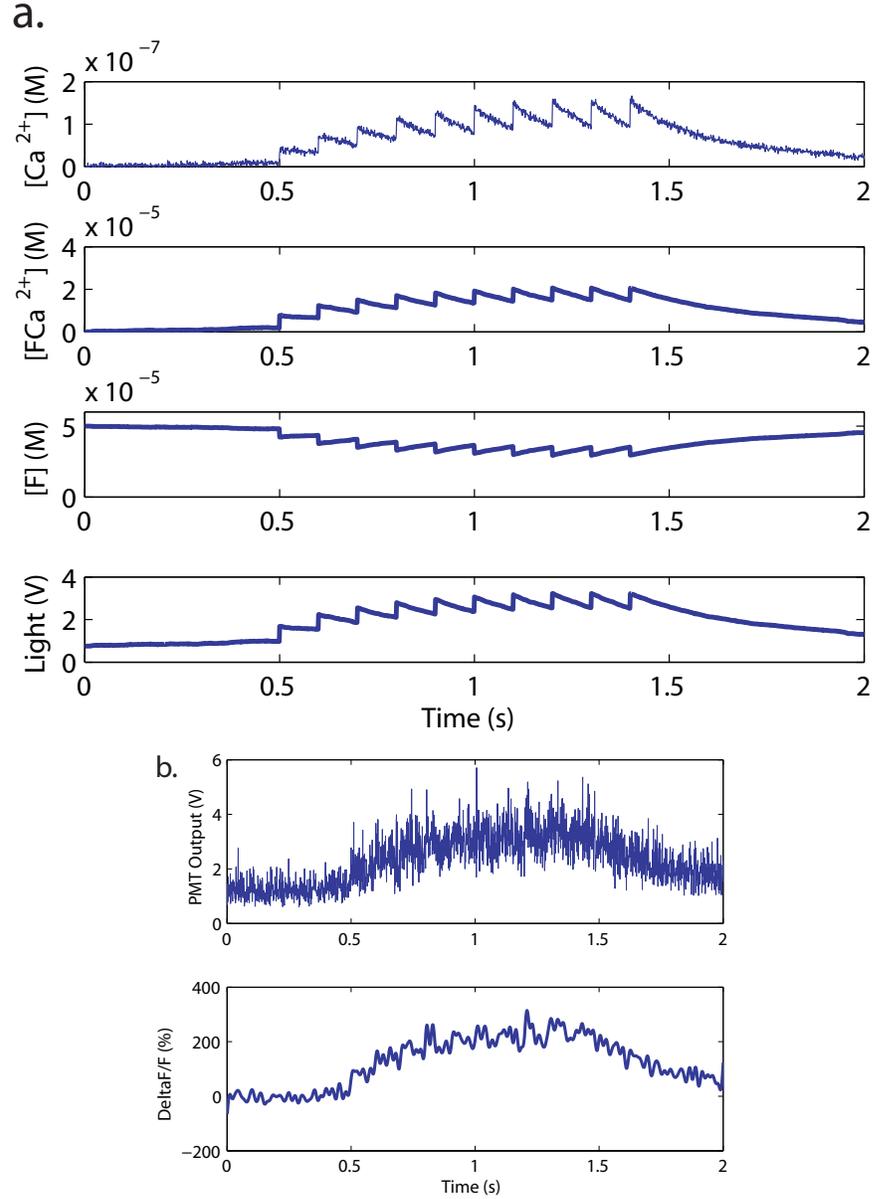


Figure 4.6: **Generation of a synthetic dataset.** (a) Plots show the values for the hidden variables of the model, generated by recursive evaluation of Equations 4.1–4.4. From top to bottom, they are free calcium, bound fluorophore + calcium, free fluorophore, and light intensity. Notice the calcium transients occurring at the times of the simulated bAPs. (b) A plot of one possible voltage trace generated by selecting random variables from the gamma distribution defined according to Equation 4.7. The bottom plot shows a low-pass filtered version of the signal normalized for baseline fluorescence, by calculation of $\frac{\Delta F}{F}$.

voltage trace is generated in the following way. For each timepoint t , calculate $A(I_t)$ and $B(I_t)$ according to Equations 4.5 and 4.6. V_t is a random variable selected from the gamma distribution with parameters A and B . This is the only observable measure of the hidden model instantiated above available to the algorithm, the results of which are described next.

4.3.1 Estimating the amplitudes of the transients

The algorithm makes use of a large number of particles, 5000 in this case, that evolve in time according to Equations 4.1–4.3. The main parameters we are trying to estimate are the amplitudes, A_j of the calcium transients induced by the bAP. At the end of iteration through the dataset, we have 5000 particles, each with an estimate of the amplitude of each A_j , and an associated likelihood of each particle.

Figure 4.7 shows the results of this algorithm in estimating the calcium transient amplitudes of the synthetic dataset generated in Section 4.3. The weight of each particle is plotted against its estimate for the transient amplitudes. Remember, the likelihood of a particular value is reflected in both the weights of the particles, and the location of the mass of particles, either a weighted histogram, or a kernel density estimate, is made from the population of particle estimates. In this case, the red curves on each graph represent a Gaussian kernel density estimate of the transient amplitudes. The vertical green line in each case shows the correct value of the amplitude of the calcium transient.

The particle filter as described in this example also estimates the total fluorophore

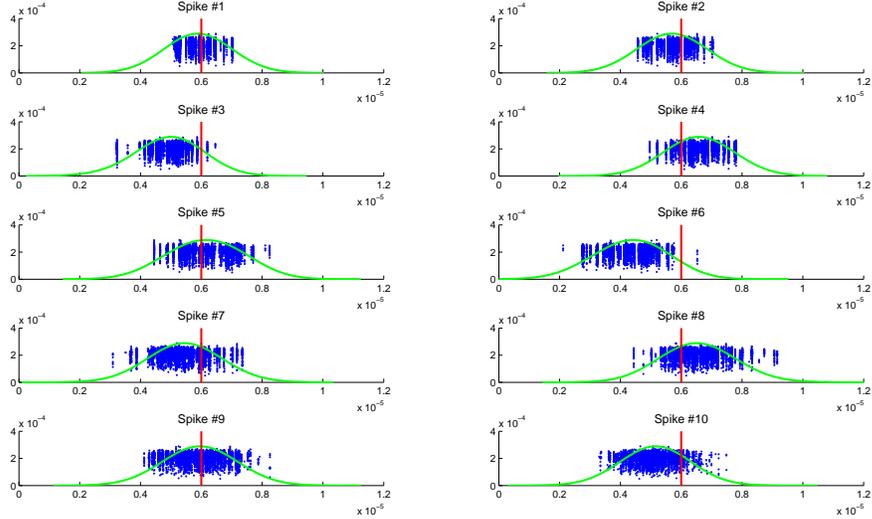


Figure 4.7: **Estimating the amplitude by SMC particle filter.** These are ten plots, representing the estimates of the ten calcium transient amplitudes from the population of particles in the particle filter. The weight of each particle is plotted against its estimate for each spike. The red curve is a kernel density estimate of the population of estimates. The vertical green lines show the correct amplitude of the underlying calcium transient.

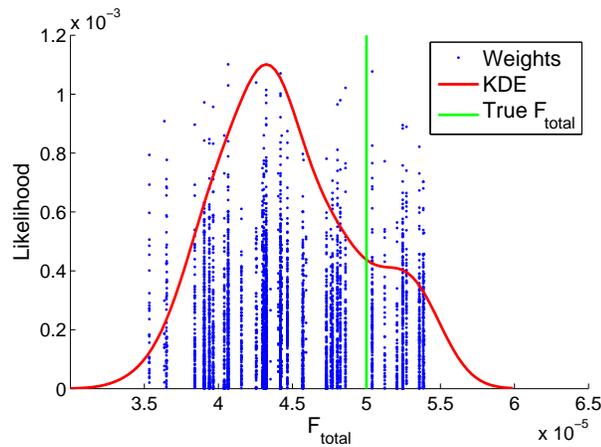


Figure 4.8: **Estimating $[F]_{total}$ by SMC particle filter.** As in the figure above, the weight of the 5000 particles is plotted against each particle's estimate, which in this case is the estimate of $[F]_{total}$. Red curve is the kernel density estimate, and the green line is the correct answer, as in the figure above.

concentration, $[F]_{total}$. Other parameters can be included as well, but do involve additional computational costs. Figure 4.8 shows the weight of each particle versus its estimate of $[F]_{total}$. Again, the red curve represents the kernel density estimate of the population of estimates. The correct answer, corresponding to $50\mu\text{M}$, is indicated as a vertical green line.

4.3.2 Generating a filtered trace

While the most important value for us to estimate is the amplitude of the calcium transients, each particle has an estimate for the values of $[FCa^{2+}]_t$, $[F]_t$, and I_t at every timestep. As such, it is possible to generate a filtered trace, to try to reconstruct the original noise-free version of our signal. This is done in a similar way as in estimating the amplitude of the calcium transients. At timestep t , each particle has an estimate of I_t . The kernel density estimate of the particle values for I_t approximates the likelihood distribution, and the peak of that kernel density estimate represents the maximum a-posteriori (MAP) estimate for the light intensity at time t .

Figure 4.9(a) shows the MAP estimate of I_t , overlaid with the true values for I_t , revealing a very good reproduction of the hidden variables by our particle filter. Remember, this estimate is made without knowledge of I_t , instead only the noise corrupted version, V_t is known to the algorithm. The raw data and the MAP estimate are shown in the bottom plot of Figure 4.9(a) for comparison.

4.3.3 Comparison to low-pass filtering

One standard technique in analyzing data obtained using fluorescent dyes is the use of a low-pass filter in an attempt to remove high frequency noise from the traces. Allow a word to justify that approach, before comparing it to modern methods, and our SMC particle filter. Given a signal, s , corrupted by a function Ψ , we measure $d = \Psi(s)$. In my view, any strategy for data analysis that improves the estimate of the signal s has merit. For example, applying function Υ to the data improves the estimate of the signal if $|\Upsilon(d) - s| < |d - s|$. By this reasoning, a low-pass filter with a reasonable cutoff frequency surely improves the estimate of the signal.

The methods presented here have attempted to better characterize how the signal is corrupted by the measurement. In other words, we are characterizing the function Ψ . Ideally, we could know Ψ exactly and invert it to calculate directly $\Psi^{-1}(d) = s$. The SMC particle filter presented here sidesteps the difficult problem of inverting the function that translates a rise in $[\text{Ca}^{2+}]_i$ into a fluorescent signal, and subsequently into a voltage at the PMT.

Comparing the performance of the particle filter directly to the low-pass filter requires a way to calculate the error of the processed trace. We calculate the error as $\sum_{t=1}^T (I_t - E_t)^2$, where T is the total number of datapoints. Here E_t is taken to mean the estimate of I_t , whether it be by MAP, or by low-pass filter. Figure 4.9(b) compares the error of the MAP estimate to the error of the low-pass filter estimate. Since each low-pass filter is defined by its cutoff frequency, the whole range must be

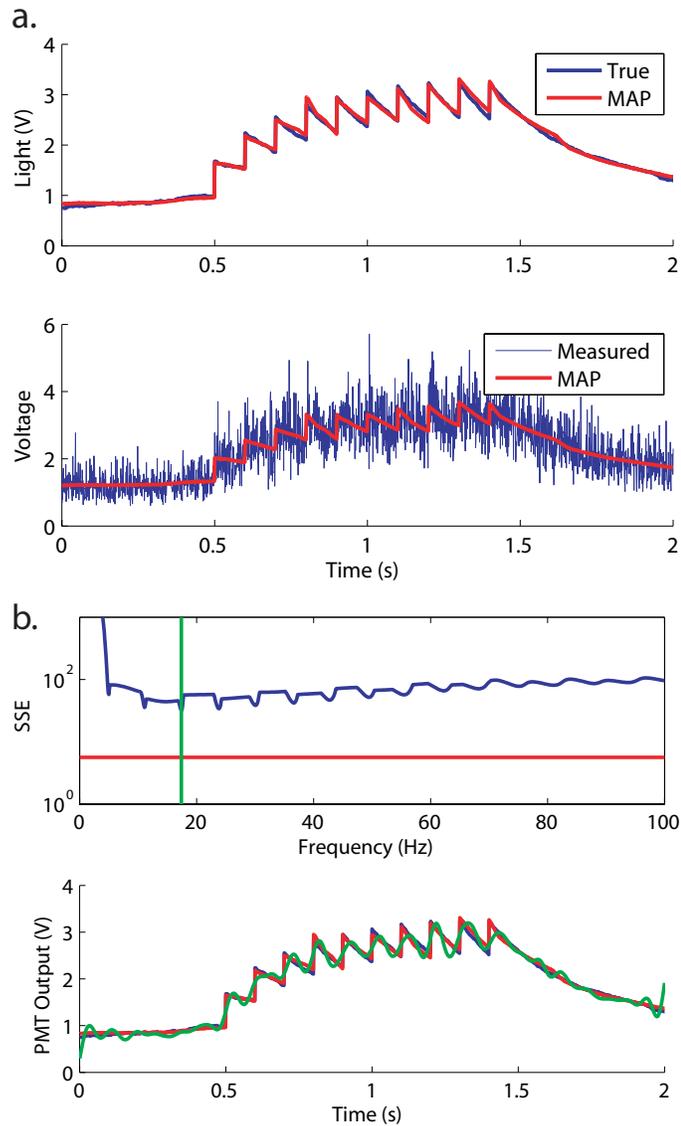


Figure 4.9: **Comparison to a digital low-pass filter.** (a) Top, plot of the true vector of I_t values (blue), overlaid with the MAP estimate of the particle filter (red). Bottom, an overlay of the MAP estimate, and the raw data, V_t , from which the MAP estimate was made. (b) Top, a comparison of the error of the MAP estimate (red line), calculated as the sum of squared differences between the estimate and the true answer, and the errors of a low-pass filter, calculated from wide range of values for the cutoff frequency. The vertical green line corresponds to the frequency of low-pass filter with the lowest error. The error of the MAP estimate is approximately an order of magnitude smaller than that of the best low-pass filter estimate. Bottom, an overlay of the true signal (blue), a low-pass filtered signal (green), and the MAP estimate (red). The frequency of low-pass filter shown here corresponds to the minimum of the error curve shown in the top plot.

tested, and as shown in Figure 4.9(b), some frequencies are more appropriate than others. As can be seen however, the MAP estimate performs much better than the low-pass filter at all frequencies.

4.4 Conclusions

I have presented here the results from my implementation of a sequential Monte Carlo particle filter designed to analyze fluorescence traces. The algorithm takes into account the way in which fluorophore binding and unbinding modifies the original calcium transient to create a signal with a fast rise time and a slow decay time, as fluorescent recordings show. Furthermore, I have characterized the voltage output of the PMT, which contains noise which is highly non-Gaussian and dependent on the intensity of the incoming light. After careful characterization of this noise profile, by measuring the distribution of voltages at many light intensities, the inverse problem, calculating the likelihood of different light intensities from a given voltage, can be done.

The results shown here demonstrate the ability to estimate the amplitudes of the calcium transients, as well as estimating the total fluorophore concentration. Furthermore, a filtered version of the signal can be generated that outperforms a digital low-pass filter, in terms of reconstructing the true signal from the noisy trace. One of the major strengths of model-based analysis is in its adaptability to the requirements of the task. The filter described here is designed to match the task of analyzing calcium fluorescence traces. As such, the filtered response (see Figure 4.9(a)) has the

capacity to rise very quickly, while decaying to baseline very slowly. This is a difficult task for filters which select for certain frequencies, such as the low-pass filter.

One drawback to the particle filter however is in the computational costs of the analysis. The task presented in Section 4.3.1, filtering a 2000 point trace with 5000 particles in MATLAB, takes about 10 minutes to run on a 1.66GHz dual core PC with 2GB RAM, whereas a low-pass filter, or the GLM presented in Section 3.1.5, takes only a few seconds. The problem becomes worse when estimating more unknown variables however, such as $[Ca^{2+}]_{base}$, or r_{ext} , since more particles are required. On the other hand, optimizations to the algorithm, and translation out of MATLAB, will likely improve the speed of the analysis by a significant degree.

Chapter 5

Discussion

5.1 Summary

Chapters 2, 3 and 4 describe my work using theoretical, experimental, and computational techniques to examine the way in which information is processed by neurons.

Using the abstract neuron model, the clusteron, we showed that learning via structural rearrangement is capable of some tasks that single neurons are not typically thought capable of. We demonstrated that not only is the clusteron capable of learning to respond preferentially to specific input patterns, but to respond to specific input patterns in a specific sequence. Furthermore, we showed they are capable of selecting common features of a set of input patterns, or under different conditions, selecting the features that distinguish those patterns.

From structural plasticity to spike-timing-dependent plasticity, our experiments examined the interaction of bAPs and EPSPs. Using calcium fluorescence, and a 3D scanning microscope, we found that pairing a bAP with presynaptic stimulation boosted the bAP associated calcium transient, most noticeably in the distal region

of the dendrites. Analyzing the data required some new strategies for data analysis, most importantly, the functional data analysis to better represent the spatial extent of the calcium transients.

Two strategies for improved measurement of the amplitudes of the calcium transients were developed, driven by the desire to make use of all the data available. The GLM model proved to be a fast and unbiased way to estimate the amplitudes of all ten action potentials in the train, with the drawback that it assumes a linear signal, and a Gaussian noise source. The SMC particle filter makes no such assumptions and was shown here to be capable of a very good estimate of the true signal underlying a synthetic dataset.

5.2 Coincidence detection

Interpreting our results further, the amplification of the calcium response to the bAP seems to support its role as a coincidence detector. When both stimuli are present, presynaptically evoked EPSPs, and a bAP, there is considerable amplification of the calcium signal. The importance of the NMDA receptor in supplying the required calcium for plasticity of the Schaffer Collaterals has been well documented (reviewed in [56]). Therefore, the role of this coincidence detector is likely to involve the guidance of the bAP towards more distal regions of the dendrite, rather than directly supplying the calcium signal for synaptic plasticity. The question remains as to why there are multiple coincidence detectors, and what is the role of each.

The spatial distribution of spike boosting is an interesting feature, since by our

data, it appears that there is not much of a change in the calcium signal proximal to the soma. Distally, however, the mechanism of coincidence detection that we see would be much more critical, since the distal region sees a much larger pairing dependent change in the signal. Furthermore, since each bAP in a train fails at a point on the dendrite closer to the soma than the preceding bAPs, the region of dendrite that experiences boosting by an EPSP changes in a complementary way. This could have implications for the plasticity rules for the regions of the dendrite depending on the firing rate of the cell. Low firing rates would yield far reaching bAPs, and the region of the cell which requires pairing induced bAP propagation would be restricted to the distal dendrites. High firing rates, on the other hand, induce bAPs which fail closer to the soma, and the region of dendrite which requires presynaptic stimulation to propagate the spike increases.

5.3 Cluster sensitivity

As mentioned in the Chapter 1, one critical piece of information that is missing from our understanding of neuronal computation is in the connectivity between individual neurons, and the subcellular location of those connections. As presented in Chapter 2, we studied a neuron model in which the spatial distribution of inputs on its dendrite changes the responsiveness of the cell. The sensitivity of the cell to different spatial patterns of input can be used as a basis for learning by rearranging the position of synapses on the tree. Learning by this mechanism predicts that connections that carry similar information would be located at nearby regions of the dendrite.

Evidence for such clustering of related inputs has been seen in at least one specialized system, namely the avian auditory system [59]. Here, the distribution of inter-synapse intervals was shown to be dependent on the nature of the input to the system.

In a recent study, however, Jia et al. [42] were able to examine the stimuli given to different regions of the cell dendrites in layer 2/3 of cortex, *in vivo*. Studying the orientation selective cells of early visual cortex, they recorded subcellular dendritic calcium signals. They showed that there were specific hotspots of orientation specific calcium influx, presumably driven by presynaptic contacts at those locations. Interestingly, most dendritic branches contained multiple hotspots specific to different orientations, suggesting that information is not relegated in a branch specific manner. This appears to contradict the predictions of the spatial learning model, which predicts that inputs which best drive the neuron, are likely to have accumulated in one region, in order to maximize their effectiveness.

One possible explanation for the contradicting results is that the signals measured by Jia et al. [42] are in fact calcium signals from clusters of synapses. The authors state that each signal they measure is likely derived from a single synapse, but only base this on the comparison to the spatial profile of single responses recorded *in vitro*. It may be true, however, that inputs that drive layer 2/3 neurons to fire are not arranged in any specific pattern on the dendrites. It is exciting to have the capabilities to observe the subcellular distribution of inputs, as this piece of information will be very enlightening as to the way neurons process information.

5.4 Model-based analysis

The Monte Carlo particle filter I present here is based on a model with some very solid assumptions, mainly that our calcium signal is the result of a reversible binding reaction with our fluorophore. Critics of model-based analysis methods point out that the quality of the results crucially depends on the assumptions made, which is a valid argument. It's absolutely true that if the model used in the analysis is wrong, then the results obtained are also wrong. The point should be made, however, that all analysis is model-based, and leaving the assumptions of the model unstated does not mean they are not made. As a simple example, estimating the mean fluorescence from a recording site can be done by calculating the sample mean of all the datapoints. The hidden model used in this approach states that the signal is completely flat, and the noise is Gaussian. We know from our measurements of the distributions of recorded voltages, with a constant input, that this assumption is wrong. As such, a better estimate of the signal can be made with assumptions that are less wrong. The maximum likelihood approach described in Section 4.1.3, is based on the actual noise distributions measured from the detector.

The strategies for data analysis presented here have proven to be very useful. Several times, after learning a new approach to a problem, I found myself as the proverbial “man with a hammer.” In particular, the GLM is a very powerful technique, which is also easily implemented. It is commonly used to analyze fMRI data [27, 41]. We use it here for measuring the amplitudes of many overlapping calcium transients. It

could be easily adapted to other common tasks, such as measuring complicated peaks of HPLC, or partially overlapping bands on western blots.

The flexibility of Monte Carlo particle filter makes it a suitable tool for numerous applications, as many researchers are currently finding. The barrier imposed by high computational costs has been partially alleviated by increasing computer speeds. Parallelization of computation, another rapidly advancing technology, will likely be a benefit in the near future as well. In the past, particle filters have been used to analyze other forms of noise corrupted time series data, like single channel recordings [9]. We use it here to make estimates of the unknown amplitude of transients which occur at known times. Another lab, however, has addressed the complementary problem, estimating the timing of action potentials [88], from a fluorescent calcium trace. Their analysis yields the most likely spike train from a fluorescence trace with transients at unknown times.

Some interesting possibilities are extending the existing models, to include a biophysically accurate description of the neuron, similar to the approach in [40]. It could be possible to estimate channel densities and kinetics as well. Perhaps an important step in the realization of that task is to optimize the experimental protocols to do so, such as in [52]. Remember that the particle filter ultimately yields the distribution of likely values for the parameter that you are estimating. The quality of the estimate depends on the width of that likelihood distribution. If the particles at the end of your analysis have a very wide range of values for a particular parameter, then confidence in the estimate is low. One cause for this could be that the experiment is not

properly designed to discriminate on that particular parameter. Thus, optimizing the experiment, perhaps through simulations, to best collapse those likelihood distribution, and thereby increase the confidence of the estimate, may be a fruitful area of research.

Bibliography

- [1] C. D. Acker and J. A. White. Roles of IA and morphology in action potential propagation in CA1 pyramidal cell dendrites. *Journal of Computational Neuroscience*, 23:201–216, Oct 2007.
- [2] J. R. Anderson, J. M. Fincham, and S. Douglass. Practice and retention: a unifying analysis. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25(5):1120–1136, Sep 1999.
- [3] B. A. Armitage and S. A. Siegelbaum. Presynaptic induction and expression of homosynaptic depression at Aplysia sensorimotor neuron synapses. *The Journal of Neuroscience*, 18(21):8770–8779, Nov 1998.
- [4] A. D. Baddeley. *Essentials of Human Memory*. Psychology Press (UK), 1999.
- [5] E. L. Bienenstock, L. N. Cooper, and P. W. Munro. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience*, 2(1):32–48, Jan 1982.
- [6] S. D. Brenowitz, A. R. Best, and W. G. Regehr. Sustained elevation of dendritic calcium evokes widespread endocannabinoid release and suppression of synapses onto cerebellar Purkinje cells. *The Journal of Neuroscience*, 26(25):6841–6850, Jun 2006.
- [7] D. V. Buonomano and M. M. Merzenich. Cortical plasticity: from synapses to maps. *Annual Review of Neuroscience*, 21:149–186, 1998.
- [8] D. B. Chklovskii, B. W. Mel, and K. Svoboda. Cortical rewiring and information storage. *Nature*, 431(7010):782–788, Oct 2004.
- [9] S. H. Chung, J. B. Moore, L. G. Xia, L. S. Premkumar, and P. W. Gage. Characterization of single channel currents using digital signal processing techniques based on Hidden Markov Models. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 329:265–285, Sep 1990.
- [10] H. T. Cline. Dendritic arbor development and synaptogenesis. *Current Opinion in Neurobiology*, 11(1):118–126, Feb 2001.

- [11] C. M. Colbert, J. C. Magee, D. A. Hoffman, and D. Johnston. Slow recovery from inactivation of Na⁺ channels underlies the activity-dependent attenuation of dendritic action potentials in hippocampal CA1 pyramidal neurons. *The Journal of Neuroscience*, 17:6512–6521, Sep 1997.
- [12] J. M. Conner, A. Culberson, C. Packowski, A. A. Chiba, and M. H. Tuszynski. Lesions of the basal forebrain cholinergic system impair task acquisition and abolish cortical plasticity associated with motor skill learning. *Neuron*, 38(5):819–829, Jun 2003.
- [13] M. Constantine-Paton and H. T. Cline. LTP and activity-dependent synaptogenesis: the more alike they are, the more different they become. *Current Opinion in Neurobiology*, 8(1):139–148, Feb 1998.
- [14] M. E. Dailey, J. Buchanan, D. E. Bergles, and S. J. Smith. Mossy fiber growth and synaptogenesis in rat hippocampal slices in vitro. *The Journal of Neuroscience*, 14(3 Pt 1):1060–1078, Mar 1994.
- [15] Y. Dan and M. M. Poo. Spike timing-dependent plasticity of neural circuits. *Neuron*, 44:23–30, Sep 2004.
- [16] C. I. De Zeeuw and C. H. Yeo. Time and tide in cerebellar memory formation. *Current Opinion in Neurobiology*, 15(6):667–674, Dec 2005.
- [17] W. Denk, J. H. Strickler, and W. W. Webb. Two-photon laser scanning fluorescence microscopy. *Science*, 248:73–76, Apr 1990.
- [18] N. L. Desmond and W. B. Levy. Changes in the numerical density of synaptic contacts with long-term potentiation in the hippocampal dentate gyrus. *The Journal of Comparative Neurology*, 253(4):466–475, Nov 1986.
- [19] M. Eberhard and P. Erne. Calcium binding to fluorescent calcium indicators: calcium green, calcium orange and calcium crimson. *Biochemical and Biophysical Research Communications*, 180:209–215, Oct 1991.
- [20] C. Eberhardt. Photomultiertube.svg. <http://en.wikipedia.org/>, Apr 2006.
- [21] F. Engert and T. Bonhoeffer. Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature*, 399(6731):66–70, May 1999.
- [22] F. Esposti, M. Ripamonti, and M. G. Signorini. Exploiting the multiplicative nature of fluoroscopic image stochastic noise to enhance calcium imaging recording quality. *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009:3589–3592, 2009.
- [23] M. F. Ethier, E. Cappelluti, and J. M. Madison. Mechanisms of interleukin-4 effects on calcium signaling in airway smooth muscle cells. *Science's STKE*, 313(1):127–133, Apr 2005.

- [24] T. Euler and W. Denk. Dendritic processing. *Current Opinion in Neurobiology*, 11(4):415–422, Aug 2001.
- [25] T. Euler, P. B. Detwiler, and W. Denk. Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature*, 418(6900):845–852, Aug 2002.
- [26] E. Fino, V. Paille, Y. Cui, T. Morera-Herrerias, J. M. Deniau, and L. Venance. Distinct coincidence detectors govern the corticostriatal spike timing-dependent plasticity. *The Journal of Physiology*, 588:3045–3062, Jul 2010.
- [27] K. J. Friston, A. P. Holmes, J. B. Poline, P. J. Grasby, S. C. Williams, R. S. Frackowiak, and R. Turner. Analysis of fMRI time-series revisited. *NeuroImage*, 2:45–53, Mar 1995.
- [28] F. Gabbiani, H. G. Krapp, C. Koch, and G. Laurent. Multiplicative computation in a visual neuron sensitive to looming. *Nature*, 420(6913):320–324, Nov 2002.
- [29] S. Gasparini, A. Losonczy, X. Chen, D. Johnston, and J. C. Magee. Associative pairing enhances action potential back-propagation in radial oblique branches of CA1 pyramidal neurons. *The Journal of Physiology*, 580:787–800, May 2007.
- [30] N. L. Golding and N. Spruston. Dendritic sodium spikes are variable triggers of axonal action potentials in hippocampal CA1 pyramidal neurons. *Neuron*, 21(5):1189–1200, Nov 1998.
- [31] A. Govindarajan, R. J. Kelleher, and S. Tonegawa. A clustered plasticity model of long-term memory engrams. *Nature Reviews Neuroscience*, 7(7):575–583, Jul 2006.
- [32] W. T. Greenough, J. R. Larson, and G. S. Withers. Effects of unilateral and bilateral training in a reaching task on dendritic branching of neurons in the rat motor-sensory forelimb cortex. *Behavioral and Neural Biology*, 44(2):301–314, Sep 1985.
- [33] R. P. Haugland. *Handbook of Biological Fluorescent Probes and Research Chemicals*. Molecular Probes, Eugene, OR, 1996.
- [34] M. Häusser and B. Mel. Dendrites: bug or feature? *Current Opinion in Neurobiology*, 13(3):372–383, Jun 2003.
- [35] D. O. Hebb. *The Organization of Behaviour*. Wiley, New York, 1949.
- [36] D. A. Hoffman, J. C. Magee, C. M. Colbert, and D. Johnston. K⁺ channel regulation of signal propagation in dendrites of hippocampal pyramidal neurons. *Nature*, 387(6636):869–875, Jun 1997.

- [37] T. Hosokawa, D. A. Rusakov, T. V. Bliss, and A. Fine. Repeated confocal imaging of individual dendritic spines in the living hippocampal slice: evidence for changes in length and orientation associated with chemically induced LTP. *The Journal of Neuroscience*, 15(8):5560–5573, Aug 1995.
- [38] D. H. Hubel, T. N. Wiesel, and S. LeVay. Plasticity of ocular dominance columns in monkey striate cortex. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 278(961):377–409, Apr 1977.
- [39] J. R. Huguenard, O. P. Hamill, and D. A. Prince. Sodium channels in dendrites of rat cortical pyramidal neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 86(7):2473–2477, Apr 1989.
- [40] Q. J. Huys and L. Paninski. Smoothing of, and parameter estimation from, noisy biophysical recordings. *PLoS Computational Biology*, 5:e1000379, May 2009.
- [41] P. Jezzard, P. M. Matthews, and S. M. Smith. *Functional MRI: An Introduction to Methods*. Oxford University Press, USA, 2001.
- [42] H. Jia, N. L. Rochefort, X. Chen, and A. Konnerth. Dendritic organization of sensory input to cortical neurons in vivo. *Nature*, 464:1307–1312, Apr 2010.
- [43] D. Johnston, J. C. Magee, C. M. Colbert, and B. R. Christie. Active properties of neuronal dendrites. *Annual Review of Neuroscience*, 19:165–186, 1996.
- [44] E. Kandell, J. Schwartz, and T. Jessel. *Principles of Neural Science, Fourth Edition*. McGraw-Hill Medical, 2000.
- [45] J. A. Kleim, T. M. Hogg, P. M. VandenBerg, N. R. Cooper, R. Bruneau, and M. Remple. Cortical synaptogenesis and motor map reorganization occur during late, but not early, phase of motor skill learning. *The Journal of Neuroscience*, 24(3):628–633, Jan 2004.
- [46] J. A. Kleim, E. Lussnig, E. R. Schwarz, T. A. Comery, and W. T. Greenough. Synaptogenesis and fos expression in the motor cortex of the adult rat after motor skill learning. *The Journal of Neuroscience*, 16(14):4529–4535, Jul 1996.
- [47] J. A. Kleim, R. A. Swain, K. A. Armstrong, R. M. Napper, T. A. Jones, and W. T. Greenough. Selective synaptic plasticity within the cerebellar cortex following complex motor skill learning. *Neurobiology of Learning and Memory*, 69(3):274–289, May 1998.
- [48] H. G. Krapp, B. Hengstenberg, and R. Hengstenberg. Dendritic structure and receptive-field organization of optic flow processing interneurons in the fly. *Journal of Neurophysiology*, 79(4):1902–1917, Apr 1998.

- [49] M. E. Larkum, K. M. Kaiser, and B. Sakmann. Calcium electrogenesis in distal apical dendrites of layer 5 pyramidal cells at a critical frequency of back-propagating action potentials. *Proceedings of the National Academy of Sciences of the United States of America*, 96(25):14600–14604, Dec 1999.
- [50] W. B. Levy and C. M. Colbert. Adaptive synaptogenesis can complement associative potentiation/depression. *Neural Network Models of Conditioning and Action: Quantitative Analyses of Behavior*, 13:53–68, 1991.
- [51] W. B. Levy and N. L. Desmond. The rules of elemental synaptic plasticity. *Synaptic Modification, Neuron Selectivity, and Nervous System Organization*, pages 105–121, 1985.
- [52] J. Lewi, R. Butera, and L. Paninski. Sequential optimal design of neurophysiology experiments. *Neural Computation*, 21:619–687, Mar 2009.
- [53] M. London and M. Häusser. Dendritic computation. *Annual Review of Neuroscience*, 28:503–532, 2005.
- [54] A. Losonczy and J. C. Magee. Integrative properties of radial oblique dendrites in hippocampal CA1 pyramidal neurons. *Neuron*, 50(2):291–307, Apr 2006.
- [55] J. C. Magee. Dendritic ih normalizes temporal summation in hippocampal CA1 neurons. *Nature Neuroscience*, 2(6):508–514, Jun 1999.
- [56] R. C. Malenka and M. F. Bear. LTP and LTD: an embarrassment of riches. *Neuron*, 44:5–21, Sep 2004.
- [57] M. Maletic-Savatic, R. Malinow, and K. Svoboda. Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity. *Science*, 283(5409):1923–1927, Mar 1999.
- [58] M. Matsuzaki, N. Honkura, G. C. R. Ellis-Davies, and H. Kasai. Structural basis of long-term potentiation in single dendritic spines. *Nature*, 429(6993):761–766, Jun 2004.
- [59] T. J. McBride, A. Rodriguez-Contreras, A. Trinh, R. Bailey, and W. M. DeBello. Learning drives differential clustering of axodendritic contacts in the barn owl auditory system. *The Journal of Neuroscience*, 28:6960–6973, Jul 2008.
- [60] B. W. Mel. The clusteron: toward a simple abstraction for a complex neuron. *Advances in Neural Information Processing Systems*, 4:35–42, 1992.
- [61] R. Mistry, S. Dennis, M. Frerking, and J. R. Mellor. Dentate gyrus granule cell firing patterns can induce mossy fiber long-term potentiation in vitro. *Hippocampus*, Jun 2010.

- [62] P. Poirazi, T. Brannon, and B. W. Mel. Pyramidal neuron as two-layer neural network. *Neuron*, 37(6):989–999, Mar 2003.
- [63] P. Poirazi and B. W. Mel. Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. *Neuron*, 29(3):779–796, Mar 2001.
- [64] A. Polsky, B. W. Mel, and J. Schiller. Computational subunits in thin dendrites of pyramidal cells. *Nature Neuroscience*, 7(6):621–627, Jun 2004.
- [65] W. Rall. Core conductor theory and cable properties of neurons. *Handbook of Physiology*, 1(part 1):39–98, 1977.
- [66] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis, Second Edition*. New York, New York: Springer Science + Business Media, Inc., 2005.
- [67] G. D. Reddy, K. Kelleher, R. Fink, and P. Saggau. Three-dimensional random access multiphoton microscopy for functional imaging of neuronal activity. *Nature Neuroscience*, 11(6):713–720, Jun 2008.
- [68] C. M. Reigeluth. In search of a better way to organize instruction: The elaboration theory. *Journal of Instructional Development*, 2(3):8–15, Mar 1979.
- [69] B. Rozsa, T. Zelles, E. S. Vizi, and B. Lendvai. Distance-dependent scaling of calcium transients evoked by backpropagating spikes and synaptic activity in dendrites of hippocampal interneurons. *The Journal of Neuroscience*, 24:661–670, Jan 2004.
- [70] M. Rubart. Two-photon microscopy of cells and tissue. *Circulation Research*, 95:1154–1166, Dec 2004.
- [71] P. Saggau. New methods and uses for fast optical scanning. *Current Opinion in Neurobiology*, 16:543–550, Oct 2006.
- [72] M. Schmidt and H. Lipson. Distilling free-form natural laws from experimental data. *Science*, 324:81–85, Apr 2009.
- [73] R. Shigemoto. Memory traces in short- and long-term cerebellar motor learning. *Annual Meeting for the Society for Neuroscience: symposium*, 2006.
- [74] S. Single and A. Borst. Dendritic integration and its role in computing image velocity. *Science*, 281(5384):1848–1850, Sep 1998.
- [75] R. E. Smith. *Psychology*, St. Paul, Minnesota: West Publishing Company, 1993.
- [76] L. Squire, D. Berg, F. Bloom, S. du Lac, and A. Ghosh. *Fundamental Neuroscience, Third Edition*. Academic Press, 2008.

- [77] A. Stepanyants, P. R. Hof, and D. B. Chklovskii. Geometry and structural plasticity of synaptic connectivity. *Neuron*, 34(2):275–288, Apr 2002.
- [78] R. J. Stokey and P. J. Lee. Approximation to the probability density at the output of a photomultiplier tube. *TDA Progress Report*, pages 42–73, Jan 1983.
- [79] G. J. Stuart and M. Hausser. Dendritic coincidence detection of EPSPs and action potentials. *Nature Neuroscience*, 4(1):63–71, Jan 2001.
- [80] G. J. Stuart and B. Sakmann. Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature*, 367:69–72, Jan 1994.
- [81] G. J. Stuart, J. Schiller, and B. Sakmann. Action potential initiation and propagation in rat neocortical pyramidal neurons. *The Journal of Physiology*, 505.3:617–632, Dec 1997.
- [82] H. H. Tan. A statistical model of the photomultiplier gain process with applications to optical pulse detection. *TDA Progress Report*, pages 42–68, Jan 1982.
- [83] N. Toni, P. A. Buchs, I. Nikonenko, C. R. Bron, and D. Muller. LTP promotes formation of multiple spine synapses between a single axon terminal and a dendrite. *Nature*, 402(6760):421–425, Nov 1999.
- [84] J. T. Trachtenberg, B. E. Chen, G. W. Knott, G. Feng, J. R. Sanes, E. Welker, and K. Svoboda. Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature*, 420(6917):788–794, Dec 2002.
- [85] M. Umekiya, M. Senda, and T. H. Murphy. Behaviour of NMDA and AMPA receptor-mediated miniature EPSCs at rat cortical neuron synapses identified by calcium imaging. *The Journal of Physiology*, 521(1):113–122, Nov 1999.
- [86] L. G. Ungerleider, J. Doyon, and A. Karni. Imaging brain plasticity during motor skill learning. *Neurobiology of Learning and Memory*, 78(3):553–564, Nov 2002.
- [87] J. J. G. van Merriënboer, R. E. Clark, and M. B. M. de Croock. Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development*, 50(2):39–61, 2002.
- [88] J. T. Vogelstein, B. O. Watson, A. M. Packer, R. Yuste, B. Jodynak, and L. Paninski. Spike inference from calcium imaging using sequential Monte Carlo methods. *Biophysical Journal*, 97:636–655, Jul 2009.
- [89] S. Watanabe, D. A. Hoffman, M. Migliore, and D. Johnston. Dendritic K⁺ channels contribute to spike-timing dependent long-term potentiation in hippocampal pyramidal neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 99:8366–8371, Jun 2002.

- [90] D. S. Wei, Y. A. Mei, A. Bagal, J. P. Kao, S. M. Thompson, and C. M. Tang. Compartmentalized and binary behavior of terminal dendrites in hippocampal pyramidal neurons. *Science*, 293(5538):2272–2275, Sep 2001.
- [91] R. E. Westenbroek, J. W. Hell, C. Warner, S. J. Dubel, T P. Snutch, and W. A. Catterall. Biochemical properties and subcellular distribution of an n-type calcium channel alpha 1 subunit. *Neuron*, 9(6):1099–1115, Dec 1992.
- [92] G. S. Withers and W. T. Greenough. Reach training selectively alters dendritic branching in subpopulations of layer II-III pyramids in rat motor-somatosensory forelimb cortex. *Neuropsychologia*, 27(1):61–69, 1989.
- [93] R. Yasuda, B. L. Sabatini, and K. Svoboda. Plasticity of calcium channels in dendritic spines. *Nature Neuroscience*, 6(9):948–955, Sep 2003.
- [94] K. Zito, V. Scheuss, G. Knott, T. Hill, and K. Svoboda. Rapid functional maturation of nascent dendritic spines. *Neuron*, 61:247–258, Jan 2009.